Senior Project

Department of Economics

The Impact of Airport Capacity Changes on the Air Quality

Karleighana Jones

Spring 2024

Advisor: *Dr. Ali Enami*

I. Abstract

The concern for the environment and how current airport industry processes contribute to emissions is growing throughout the United States. The air quality index is a great measure of the environmental conditions and connects to the health of the people within the counties. With airport capacity and enplanement levels rising between 1999 and 2019, the number of individuals using airplanes and the frequency of take-off and landing procedures are both increasing. Optimizing internal operations, weighing the economic benefits, and determining the residential impacts are avenues previous studies have taken to measure the impact airports have on the surrounding communities. As the airline transportation industry expands, other researchers have focused on its impact on air quality and the emission particles in residential areas. Despite the extensive results garnered from the previous studies, there has yet to be an analysis and discussion about the varying impacts on the air quality for different sizes of the airports across the nation with such longevity. In other words, this study is unique with its widespread 20 years of data, the analysis done at the county level, and the relationship between the distinction with the size of airports within a county with good air quality days. At the conclusion of the analysis, the statistically significant results demonstrate the overall impact of airport capacity expansion depends on the capacity level as there are both positive and negative relationships present throughout the data.

II. Table of Contents

III. Introduction

The ongoing issue with air quality changes in the United States are a result of a combination of factors. These factors include direct and indirect human contributions such as burning fossil fuels, changing weather conditions, and wildfires. An overwhelming abundance of air pollution introduced into the atmosphere has a negative impact on the air quality and the rate of global warming. Even technological advances that revolutionized progress have an influence on air quality.

One of the factors that impact air quality are the technological advances that modernize methods of transportation, specifically air travel. The airline industry contributes mass amounts of carbon dioxide and other pollutants into the atmosphere with the burning of fossil fuels to power aircrafts. The utilization of airplanes as a means of transportation has a lasting impact on the environment.

This research paper aims to determine the overall impact the expansion of airport capacity has on air quality in the surrounding counties from 1999 to 2019. How much is the introduction of more flight departures through an airport's expansion, changes in hours of operation, or the increase in airport use affecting the air quality in the surrounding cities?

With the continuous emissions of pollutants into the atmosphere, the risks of health issues among the population are constantly put in jeopardy such as the rise in cancer, respiratory infections, and mortality rates. The societal consequences of the rise in pollutants in the atmosphere are connected to the issues from the rise in global warming. In a generation that is environmentally cautious, acknowledging the eminent dangers from current societal practices and how it threatens the luxury of clean air is the first step.

The motivation behind the research is that poor air quality negatively affects the health of individuals, and the airline industry is a contributing factor of the worsening air quality. The continuous emissions of pollutants are contributing to the rate of global warming. For these reasons, understanding the inner workings of airport's internal operations and the impact capacity changes made over time is crucial since it relates to the future and health of mankind.

Most of the previous research papers discuss airline emissions and pollutant particles emitted into the residential areas for a handful of airports within a period of a few years. They measure the data collected within a finite time period, typically less than 5 years. This research focuses on the changes in airport capacity in different airports across the nation, rather than just one, and utilizes the air quality data in those regions to measure the overall impact these changes have on the county air quality. The enplanement data is used as an indirect measure of airport emissions as an increase in capacity levels is correlated with an increase in pollutant emissions from the airplanes.

This study uses a Log-Log Two Way Fixed Effects to detail the relationship between county air quality levels and enplanement capacity, specifically for good air quality days. The main results from this analysis suggests that the overall effect of airport expansion is dependent on the capacity. The statistically significant results highlight the airport size stipulation as there are both positive and negative relationships present.

The remainder of the paper is organized as follows: the Literature Review details previous research papers discussing air quality, airports' contribution to air quality, and benefitting factors as a result of airport expansions; the Data section introduces the main databases utilized for further analysis and preliminary summary statistics; the Theory and

Methodology dives into the economic theory, expected results, and the structure of the models; the Results section; the Conclusion section; followed by References and the Appendix.

IV. Literature Review

Technological and industrial advances worldwide that produce air pollutants negatively impact the surrounding air quality. In turn, it deteriorates the health conditions of individuals within the vicinity of these expansions. Countries seeking to take advantage of economic opportunities must weigh the impact of the proposed expansion on the community, the environment, and the health of its residents prior to pursuing them. Large processing industries that require multiple avenues of operations working in tandem to be successful, such as factories and airline industries, have a larger overall impact on the environment and respective air quality. The desire to expand industrial sectors vary by country, yet the common thread is the potential economic benefits as a result from the growth. Apart from airport expansions influencing air quality levels, factory expansions have similar impacts on the air quality and the health of individuals in close proximity.

Industrial sector factory expansion deteriorates the air quality, which has adverse health effects. Factory emissions raise the prevalence and frequency of different diseases in the surrounding communities. For instance, Beketie et al. (2021) find that the expansion of the industrial sector in Ethiopia resulted in the production of cement factories that simultaneously stimulate the economy and deteriorate the health of the residents due to the air pollution from the extensive energy consumption required to fuel the cement factories. Children and adults living in the vicinity were negatively affected by respiratory diseases as 62.9% experienced chronic respiratory symptoms with lasting impacts for generations (Beketie et al., 2021). Apart from Ethiopia’s industrial expansion, between 2007 and 2016 in over 300 China cities, the industrial land expansion worsened air pollution as the rapid changes in land-use altered the regional landscape and environment (Li et al., 2022). In both cases, the expansion of the industrial sectors and the changes in how the respective lands were utilized resulted in worsening levels of the surrounding air quality. The airline industry also impacts surrounding air quality.

Residents and the surrounding housing prices are impacted by the noise, air, and water pollution from airports. The daily operations needed for airports to function properly contribute to the worsening environmental effects. Such activities include the operation of the aircrafts, the ground service equipment, cleaning, maintenance, fuel storage, and the presence of new construction. Each of the processes involved with the utilization of airplanes as frequent modes of transportation influence the water, air, and noise quality in the surrounding areas (Luther, 2007). A quasi-experimental study from before and after the relocation of the Hong Kong Kai Tak Airport in 1998 show that by diminishing the noise pollution from airports, the surrounding residential communities experience a rise in housing prices by 24.43% relative to the control group (Zheng et al., 2020, para. 39). Another study determines the damages that residents within a geographical area are willing to receive on average \$100-\$400 per person annually given their proximity to the airport (Wolfe et al., 2014). These research papers allude to the negative relationship between residents' experience and their proximity to an airport producing noise pollution. After the Hong Kong airport relocation, housing prices that were now further away from the airport increased. These studies emphasize the prevalence of airline pollution, the adverse effects on the residents within earshot, and the proposed monetary incentive most residents would appreciate receiving for the hardships endured. The airport pollution can be broken down by the internal operations to determine which contributes the most.

At airports, the highest emission distribution from internal operations is during the take-off and landing cycles. Increasing the frequency of the airport in- and out-bound flights due to capacity expansion would further contribute to the worsening air quality. A study of the take-off at LAX airport in 2005 and 2006 focused on the particles emitted from the airplanes on and around the runway. After taking into account the wind direction, the researchers found the highest spikes in carbon emissions was correlated with the take-off times from LAX (Zhu et al., 2011). Their results detail that "aircraft takeoffs contributed 53.5% to the total UFP concentration at the blast fence of 25R runway, other airport operations contributed 45.8% and the background ambient accounted for 0.7%," (Zhu et al., 2011, para. 23). In other words, increasing the airport capacity elevated the number of flights taking off, which increases the output of airplane ultrafine particle (UFP) emissions in the surrounding areas. Additionally, air traffic data in 2018 from Turkey's largest commercial airline company focused on determining the emissions levels of different aircrafts during the landing and take-off cycles (Ekici and Sevinc, 2021). The results separate the breadth of the implications based on aircraft type, the stage of the flight process, and the emission particle being measured. The researchers found that the maximum emissions per landing and take-off cycle happens during take-off as the airplanes utilize the most fuel. They suggest a potential solution that considers the environmental impacts as well as the needs for the airports to expand capacities; that through choosing optimal aircraft-engine combinations with the least emissions, the landing and take-off cycles may have the ability to increase, as the unique airplane configuration reduces the environmental impact.

Airports can reduce emissions through streamlining internal operations; specifically timely taxiing and considering the runway options. The airport and airline conditions that determine which runways to take-off from impact the air quality in surrounding areas. In the residential communities surrounding The General Edward Lawrence Logan International Airport near Boston, runway configuration is highly linked to pollutant concentrations and flight activity with regards to preferred runways and winds carrying these pollutants (Hudda et al., 2020).

Hudda et al. (2020) continues by stating that when the observed residence was downwind of the airports, there was a higher concentration of pollutants; also, overhead landing operations has more particle pollutant emissions compared to the closest runway take-offs. This suggests that the decisions regarding the maneuvering of aircrafts on a chosen runway and which runway airplanes are directed to land on are more than simply part of the flight operation. Along with the contributing emissions from the landing and take-off cycles, the choice of runway at airports proves to be a significant contributor to the air pollution. These procedures combined with the weather and wind patterns influence where the pollutants end up in the surrounding neighborhoods.

Additionally, certain measures could be taken to minimize the environmental pollution during taxi and take-off operations at airports by finding the most cost-effective adjustments to pushback control and thrust settings. The research conducted in 2007 at the Detroit Metropolitan Wayne County Airport sought to determine the optimal pushback control to reduce congestion leading up to the runway and thrust levels to produce as little pollution as possible while not compromising the functionality of the aircraft. The researchers find that the strategies are arguably effective with altering the trust setting between 75%-81%; the resulting total fuel combustion-related costs and environmental costs are minimized within that range (Ashok et al., 2017). This argues that adjusting the current levels at which pilots and airline professionals operate airplanes with regards to mindful departure from the gate and levels of thrust at take-off can bode significant changes in emissions into the environment. A few other alterations that can reduce emissions include using biofuels, direct routing, or by improving the calculations used to derive the cost index for airlines (Edwards et al., 2016). Given the active role of the Environmental Protection Agency and other governmental agencies whose purpose is to regulate and minimize the pollution, planning for future airport capacity expansions depends on the current capacity, location, and popularity of the airports. By creating a cost-benefit analysis, executives and lead officials can determine the necessary measures needed to regulate environmental impacts from the anticipated growth of the airport.

Airport expansions have positive economic benefits with the creation of jobs and income benefits, despite the extensive airport noise and air pollution. In 2018, the five busiest airports generated $181.4 billion of direct economic benefits to their respective communities and regions (Nataraja and Peterson, 2019). When weighing the economic impact from airport usage to society, it complicates the issue with air quality due to the monetary significance and wealth airport expansion would generate for the residents in the surrounding communities. Results from the airport capacity changes within the Chicago region in 1993 forecasted an exponential increase in employment projections (Hewings et al., 1997). These included increased employment opportunities in agriculture, construction, trade, manufacturing, trade and more industries compared to if there were airport capacity limitations. In other words, these studies provide an economic argument for the value airports bring to surrounding areas and related industries with the potential employment and income benefits that come from its expansion.

Most literature reviewed discuss studies containing airline emission data along with the presence of pollutant particles in the surrounding residential vicinity. Rather than utilizing emission data from airports, the variable of interest is the volume changes in enplanements or departures from the airports. The contribution this research aims to achieve is the analysis of tangible changes in the capacity for airports across the U.S. and how it relates to the categorization of types of air quality days. Through this approach, the enplanements behave as an indirect measure of flight data to consider the correlation to the respective air quality data.

V. Data

The main databases used in this research contain annual air quality data by county and enplanement airport data by airport. The first database contains data from counties in the United States with annual air quality classification days from the United States Environmental Protection Agency from 1999 to 2019. To create a consistent measure of air quality that is comparable across counties and over time, the total number of days with AQI measurement is used to scale the underlying variables[1]. The enplanement data for airports in the U.S. is from the Federal Aviation Administration Database. The layout for each annual report varies, so further cleaning was needed to combine them all. In order to connect the two main databases, the airport database needs to be at the county-level to match the county-level air quality data. A third database that connects the cities to their respective counties from The United States Cities Database creates a crosswalk between the two main databases to enable further analysis toward addressing the research question. As airports expand their capacity levels, the number of planes contributing to the air quality would arguably increase, causing the air quality levels to worsen with rising AQI, as the higher the AQI the worse type of day it would be categorized.

Table 1 details the descriptive statistics, such as mean, minimum, maximum, and standard deviation, of each type of day out of the total number of days where AQI was measured as well as enplanement level data at the county level between 1999 and 2019. It is important to note that about 97% of the days measuring AQI are good or moderate days and that the average enplanement level is around 1 million individuals, with a maximum capacity of 53.5 million passengers in a given county and year.

---

[1] Specifically, the six new variables created are Good Days (GD), Moderate Days (MD), Unhealthy for Sensitive Groups Days (USGD), Unhealthy Days (UD), Very Unhealthy Days (VUD), and Hazardous Days (HD).

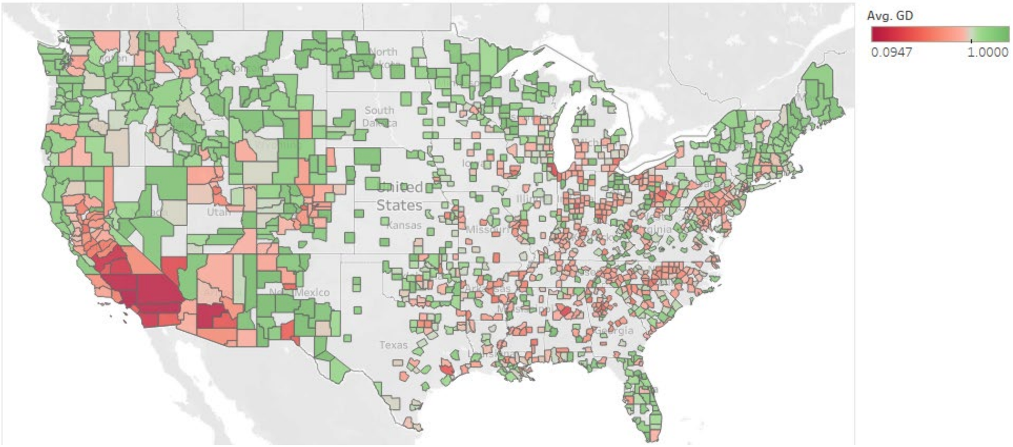
Table 1: Summary Statistics for AQI Day Types and Enplanement Levels from 1999 to 2019

Main Variables	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Good Days	22,394	74.45%	18.27%	1.37%	100.00%
Moderate Days	22,394	22.55%	15.41%	0.00%	92.62%
Unhealthy for Sensitive Groups Days	22,394	2.44%	3.91%	0.00%	46.58%
Unhealthy Days	22,394	0.51%	1.62%	0.00%	31.23%
Very Unhealthy Days	22,394	0.05%	0.35%	0.00%	13.97%
Hazardous Days	22,394	0.01%	0.18%	0.00%	10.14%
Enplanement Level	13,025	3,094,306.48	7,797,286.59	10,003	53,515,982
Population	63,879	97,697.32	316,958.19	55	10,123,521
Income	63,879	33,502.46	11,328.76	8,978	260,038

Sources: EPS (2024), FAA (2024), and own calculations.
Notes: The type of days (e.g., Good Days) is constructed by dividing the corresponding number of days in a year by the total number of days AQI was measured that year. Enplanement level data is aggregated based on the county in which the airport is located. Air Quality Data and Enplanement Data is at the county level from 1999 to 2019.

Figure 1 depicts the ratio of good air quality day distribution based on the county between 1999 and 2019. The color red corresponds to fewer percentages of good air quality days and green counties signify, on average, a larger portion of the days where AQI was measured was categorized as a day with good air quality. One can gather that counties that contain large metropolitan areas, including Atlanta and Chicago, have fewer good air quality days over the twenty-year time period.

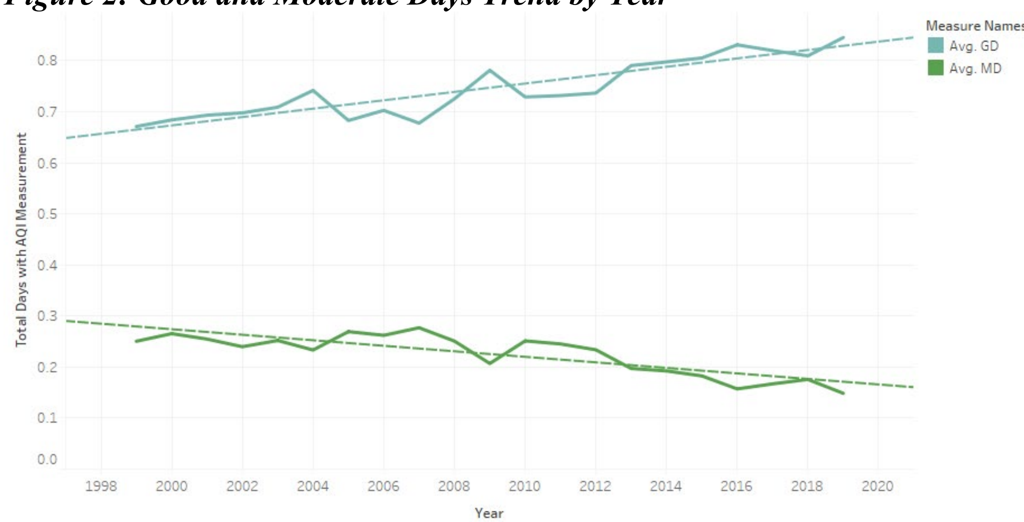
Figure 1: Average Ratio of Good AQI Days Measured by County



Source: EPA (2024) & OpenStreetMap by Mapbox (2024), with own calculations.
Notes: Map based on AQI Good Day Types (GD) at the county level from 1999-2019. Color red corresponds to fewer percentage of good days and green corresponds to a greater percentage of good air quality days over 20 years.

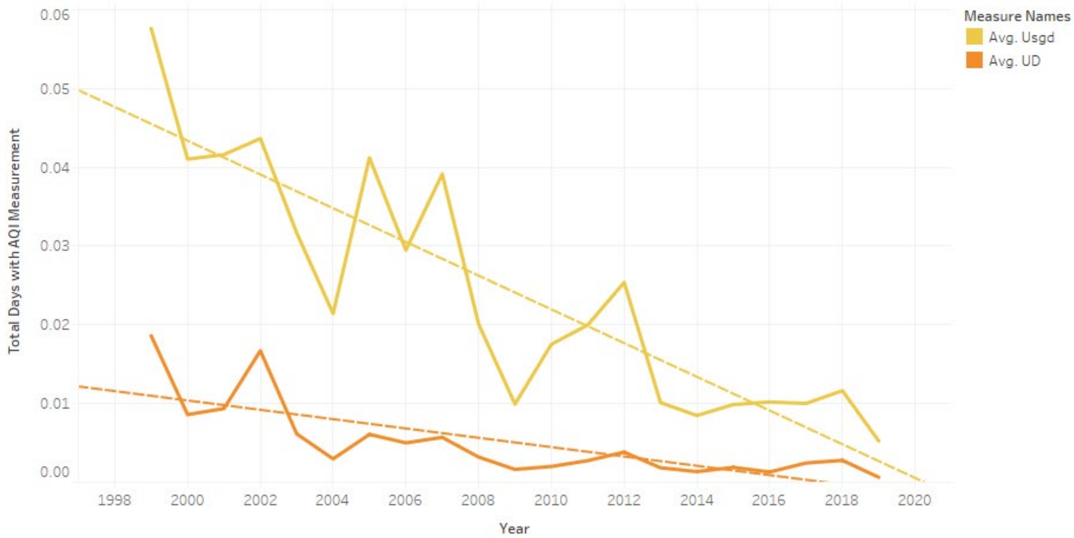
Figures 2, 3, & *4* demonstrate the six trends of the average type of day. A key finding is that the ratio of good air quality days has a positive trend whereas the moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, and hazardous days all have a negative trend. At a glance, the initial air quality trends are promising, as the percentage of good air quality days are rising steadily and the non-good days are decreasing in the same period.

Figure 2: Good and Moderate Days Trend by Year



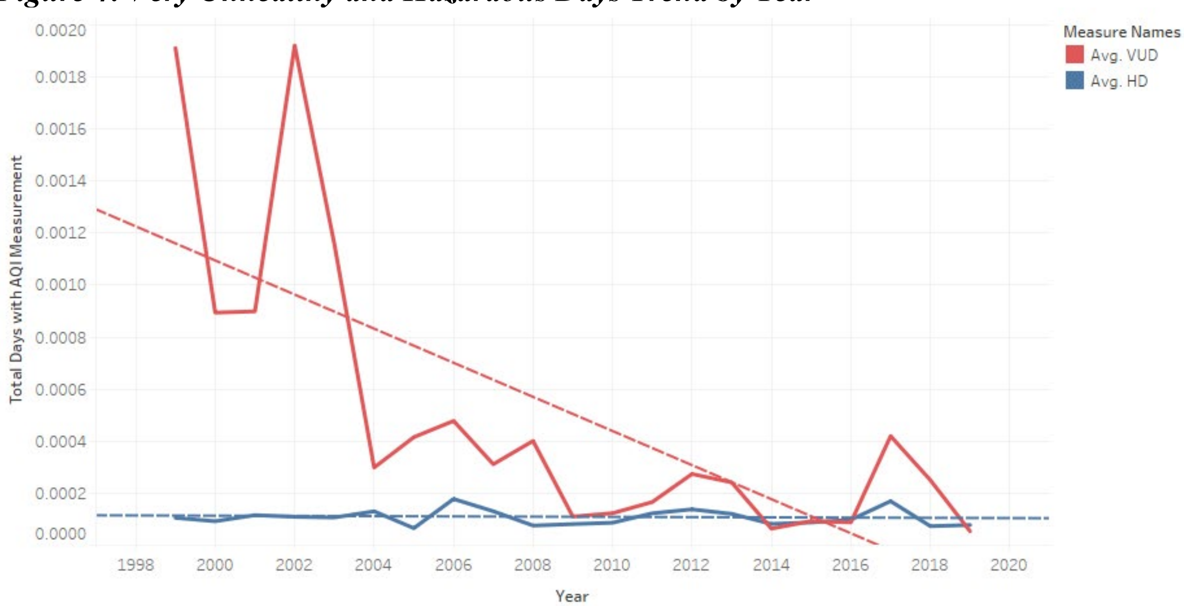
Source: EPA (2024) with own calculations.
Notes: The type of days (e.g., Good & Moderate Days) is constructed by dividing the corresponding number of days in a year by the total number of days AQI was measured that year. Air Quality Data is at the county level from 1999 to 2019.

Figure 3: Unhealthy for Sensitive Groups and Unhealthy Days Trend by Year



Source: EPA (2024) with own calculations.
Notes: The type of days (e.g., Unhealthy for Sensitive Group & Unhealthy Days) is constructed by dividing the corresponding number of days in a year by the total number of days AQI was measured that year. Air Quality Data is at the county level from 1999 to 2019.

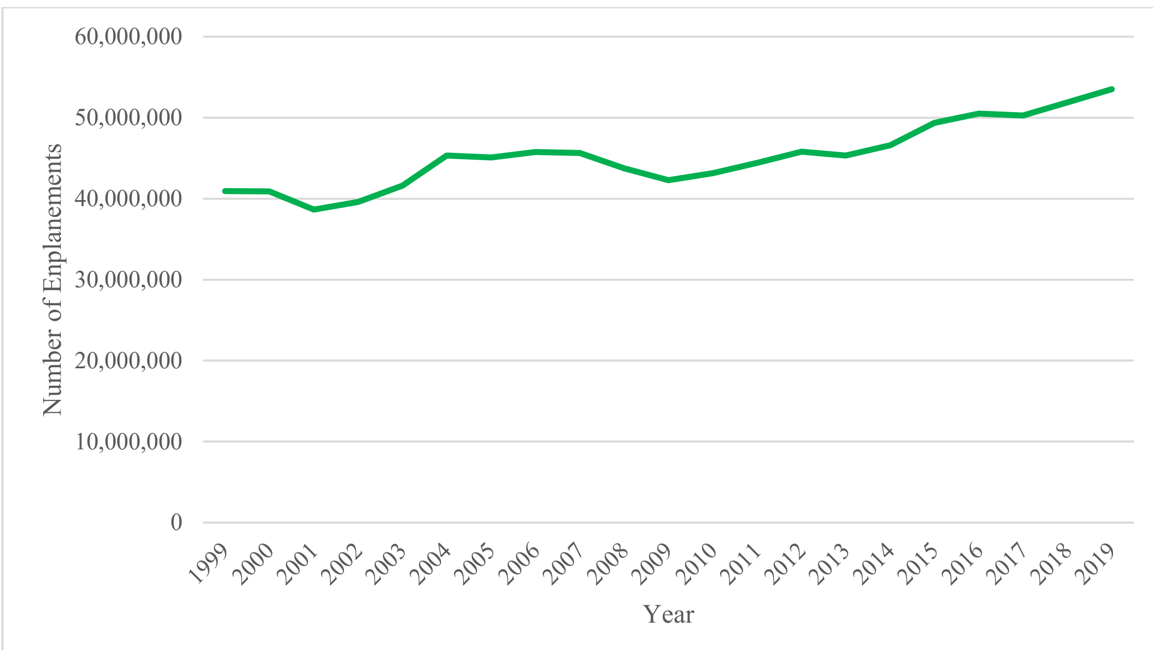
Figure 4: Very Unhealthy and Hazardous Days Trend by Year



Source: EPA (2024) with own calculations.
Notes: The type of days (e.g., Very Unhealthy & Hazardous Days) is constructed by dividing the corresponding number of days in a year by the total number of days AQI was measured that year. Air Quality Data is at the county level from 1999 to 2019.

Figure 5 demonstrates that the highest number of enplanements in a given county is increasing between 1999 and 2019[2]. The United States population[3] during the same period increases by about 50 million individuals (FRED, 2023). Naturally, the highest level of enplanements in a given county in a given year would increase within the twenty-year period since the total population also experiences an increase in the same period.

Figure 5: Highest Number of Enplanements in a One County per Year (1999-2019)



Source: Federal Aviation Administration (2024) with own calculations.
Notes: Enplanement data is aggregated based on the county in which the airport is located. The highest enplanement capacity per year in a given county is represented by the green line.

[2] Each airport is associated with their city location in the FAA Database (2024). By using the U.S. Cities Database (2024), it assigns county names with each airport based on the corresponding cities each airport resides in. The enplanement data is then aggregated based on the county in which the airport is located to create summary statistics.
[3] The average annual total population changed from about 280 million to 330 million between 1999-2019, as seen from FRED (2023).

VI. Theory and Methodology

A. Theory

Does airport capacity expansion truly influence the air quality levels in the counties across the United States; or are the decisions surrounding the expansion of the airports environmentally conscious enough to off-set what would be creating worse air quality levels? The expected effect of airport expansion on air quality is negative. Airports seek to maximize utility through increased capacity levels. The intricate process in which airport executives approve the capacity expansions have the ability to consider the opportunity whether to adjust plans due to a concern for the potential negative air quality costs or to proceed regardless of environmental implications or societal backlash. Despite the overall trends of the ratio of good days and enplanements increasing over the twenty-year period, the summary statistics fail to demonstrate the potential negative relationship at the county level.

Thus, the true implications of this relationship can be discussed when determining the "total cost" for an airport to expand its capacity. Everything else constant and in the absence of any laws or public pressure, airport expansion operates without concern of its effect on air quality as it is cheaper to expand an airport without caring for air quality. However, if there are environmental laws in place and active public pressure that subsequentially increases the "total cost" of expansion, then there is an economic reason for the owner of a public or private airport to allocate more funds to invest in an environmentally-friendly expansion to reduce the fines and costs related to public pressure. In conjunction with the construction costs, strategically planning ahead and incorporating environmentally-conscious plans may end up with airport owners having a "total cost" that is less than if they fail to consider the environmental impacts upfront. Also, another theory is that the increase in capacity for smaller airports may be too small to have a significant impact on the overall air quality in the county; the effect could go undetected. On the other hand, the airports in smaller counties may end up worsening the surrounding air quality as they may "get away" with operations that lack environmental considerations.

The difference in "total cost" determines whether the airport expansion will be environmentally conscious. The theory is that the airport owner's "total costs" play an important role in whether an expansion is environmentally friendly or not. The owner must weigh whether it is more expensive to incorporate an environmentally conscious plan to take into account the adverse effects and adjust accordingly, or be ignorant of the potential societal and environmental impact with the expansion and then suffer the monetary and perceptual consequences. These environmentally-conscious considerations could include changes to internal operations contributing to pollution emissions at airports such as the type of fuel used, timely taxiing, adjusting thrust levels, thoughtful runway choice, and more. The airport capacity expansion increases the day-to-day revenue with the ability to accommodate more customers, yet it incurs additional indirect consequences, which is part of the "total cost" of expansion, with the increasing modes of transportation necessary to get individuals to the airports.

These considerations are related to the "minimizing total cost" theory, as the airport expansion effects can also vary based on the size of the expansion and the current usage of the airport. In other words, the expansion of large-scale operations surrounding the Hartsfield–Jackson Atlanta International Airport in Georgia is more likely to consider incorporating an environmentally conscious strategic plan in their operations to satisfy the public's concern for the pollution emissions from such a bustling and well-known airport. However, smaller-scale airports may have more incentive to cut corners to reduce their initial "total cost" of expansion due to lack of sufficient funding to incorporate the environmental concerns or their lack of a prominent airport presence so public pressures may be minor. In this case, small expansions may hurt the air quality more than big expansions.

It is important to consider opposing theories to understand the potential impact the airport capacity expansion has on the air quality levels in the surrounding counties. Therefore, the hypothesis for the results of this research is that the relationship between the airport enplanement levels and the AQI depends on location and size of the airports.

B. *Empirical Methodology*

The analysis of the relationship between county air quality levels and enplanement capacity is through the use of Log-Log Two-Way Fixed Effects Models. This type of model, with county and year fixed effects, aid in controlling the potential bias during the analysis. The base linear regression equation for the econometric models is as follows:

$$lnAQI_{ti} = B_0 + B_1 lnEnplanement_{it} + X_{it} + B_2 County_i + B_3 Year_t + \varepsilon_{it} \tag{1}$$

$lnAQI_{ti}$ is the indicator of air quality measured in county *i* at time *t*. This indicator is a specific type of day; in this study, simply Good Days. This base model could be utilized to estimated separately for each type of day; Good Days, Moderate Days, Unhealthy for Sensitive Groups Days, Unhealthy Days, Very Unhealthy Days, and Hazardous Days, as defined previously in the data section.

$lnEnplanement_{it}$ measures the aggregated airport capacity levels of all the airports within a given county *i* at time *t*. X_{it} represents a few control variables in the analysis, including population changes and personal incomes. The variables make up factors separate from the airport enplanement data that arguably contribute to the change in air quality over time. By including these control variables, the main coefficients are better suited to represent causal effects and not merely some correlation between two variables because of missing variables in the models. *County* and *Year* are county and year fixed effects, respectively. Lastly, ε_{it} is the white noise.

VII. Results

Table 2 depicts the results when separating the aggregated enplanement data into quartiles and analyzing the impact on the ratio of good air quality days without control variables. Quartile 1 has between 10,000 and 92,316 enplanements within a county; Quartile 2 contains enplanements between 92,316 to 429,523; Quartile 3 has enplanements ranging from 429,523 to 3,416,745; and Quartile 4 has between 3,416,745 and 53,515,982 enplanements. The final column demonstrates the overall relationship when taking into account All Counties with good air quality and enplanement data.

Table 2: Log-Log Two-Way-Fixed-Effects Approach for Quartiles 1 through 4

Regressors	**Quartile 1**	**Quartile 2**	**Quartile 3**	**Quartile 4**	**All Counties**
Ln (Enplanement)	-0.01493 (0.01171)	-0.10467*** (0.01829)	0.10442*** (0.03421)	-0.00760 (0.02866)	-0.02190*** (0.00754)
Intercept	-0.10987 (0.11529)	0.98069*** (0.22034)	-1.92429*** (0.48666)	-0.60417 (0.45288)	-0.16052 (0.11524)
County Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Number of Observations	2,160	2,160	2,160	2,158	8,638
Adjusted R-Square	0.7187	0.8458	0.8381	0.9111	0.8763
Overall Significance	1.22E+02***	1.13E+02***	1.30E+02***	4.72E+02***	1.85E+02***

Source: EPA (2024), FAA (2024), with own calculations.

Notes: Robust standard Errors are in Parentheses. *, **, *** indicate 10%, 5%, and 1% significance levels, respectively and are clustered at county levels. The data is unbalanced as the number of counties in each state in a given year are not constant over time, as counties are added to the data or begin measuring AQI. Quartile 1 = Model 1 tec. See the text above table for information about enplanement cut off values for different quartiles.

As seen in *Table 2*, if enplanements in a county within Q2, between 92,316 and 429,523 enplanements, go up by 1%, then the ratio of good air quality days goes down by about 0.10%.

For Q3, if enplanements in a county increase by 1%, it is associated with an increase in the ratio of good air quality days by about 0.10%. Both effects are statistically significant at 1%.

Another statistically significant coefficient is for All Counties, which suggests that if enplanements in any county increases by 1%, it is associated with a decrease in the ratio of good air quality days by about 0.02% at a 1% significance level. In other words, the overarching impact is slightly negative across the United States.

For the smaller county enplanement capacity levels, an increase in enplanements is correlated with a decrease in the ratio of good days. Their operations lack the funding of larger airports and they lack the sufficient resources to have flexibility in their budget and operating expenses to stay below the accepted threshold. Smaller airports also have the luxury of more anonymity than more well-known airports in other counties.

For the models that lack significance in their respective enplanement coefficients, the explanation could be their enplanement levels range is too wide to detect the discrepancies between the changes in enplanement. Thus, further division of the data is needed to potentially discover these insights.

Table 3 demonstrates the aggregated enplanement analysis without control variables separated into deciles (10% increments) at the county level from lowest to highest[4].

[4] Deciles 1-10 reflect the range of enplanements in a given county in 10 percentile increments. Decile 1 reflects the 10th percentile between 10,000 and 23,240 enplanements within a county; Decile 2 represents between 23,240 and 54,728; Decile 3 shows between 54,728 and 138,423; Decile 4 is between 138,423 and 263,217; Decile 5 represents between 263,217 and 429,523; Decile 6 is from 429,523 and 846,369; Decile 7 is between 846,369 and 2,128,405; Decile 8 reflects between 2,128,405 and 5,054,191; Decile 9 is between 5,054,191 and 17,274,506; Decile 10 shows between 17,274,506 and 53,515,982 enplanements.

Table 3: Log-Log Two-Way-Fixed-Effects Approach for 10 Deciles

Regressors	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
Ln (Enplanement)	0.03199 (0.03279)	-0.03109 (0.03061)	-0.10884*** (0.03740)	-0.07416* (0.03836)	-0.04910 (0.06151)	0.26241*** (0.09406)	-0.03924 (0.05229)	0.01719 (0.06751)	0.12463** (0.06254)	-0.24446 (0.15604)
Intercept	-0.55835* (0.31877)	0.24072 (0.31116)	1.02385** (0.44256)	0.69008 (0.46486)	0.19881 (0.76371)	-3.73440*** (1.23934)	0.07365 (0.75029)	-0.96502 (1.02672)	-2.56658*** (0.97930)	3.19453 (2.57975)
County	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	864	864	864	864	863	864	864	864	864	863
Adjusted R-Square	0.8004	0.7217	0.8732	0.8729	0.6651	0.7717	0.8773	0.8727	0.9145	0.9082
Overall Significance	1.37E+02***	5.00E+02***	4.17E+02***	4.21E+03***	1.54E+02***	2.43E+03***	1.17E+02***	2.57E+02***	5.94E+03***	4.74E+02***

Source: EPA (2024), FAA (2024), with own calculations.

Notes: Robust standard Errors are in Parentheses. *, **, *** indicate 10%, 5%, and 1% significance levels, respectively and are clustered at county levels. The data is unbalanced as the number of counties in each state in a given year are not constant over time, as counties are added to the data or begin measuring AQI. Split into 10 Deciles between 10,000 to 53,515,982 enplanements; 1st Decile is Model 1, going up the decile ranges and 10th Decile is Model 10.

Table 3 results reflect three statistically significant findings in Deciles 3, 4, 6 and 9. If we assume the number of days where AQI was measured doesn't change much, in other words it stays around 365 days or 1 year, then the only change is the actual change in number of total good air quality days. Decile 3 shows that if a county with 54,728 to 138,423 enplanements goes up by 1%, then the ratio of good air quality days decreases by 0.109% at a 1% significance level. For Decile 4, if a county between 138,423 and 263,217 enplanements goes up by 1%, then the ratio of good air quality days goes down by 0.074% at a 10% significance level. For Decile 6 in *Table 3*, if a county between 429,523 and 846,369 enplanements goes up by 1%, then the ratio of good air quality days goes up by 0.262% at a 1% significance level. Decile 9 contains the final statistically significant coefficient; if enplanements in a county between 2,128,405 and 5,054,191 enplanements go up by 1%, then the ratio of good air quality days goes up by 0.125% at a 5% significance level. For the models that lack significance in their respective enplanement coefficients, one possible explanation could be their enplanement levels are in between the positively and negatively significant models' enplanement levels. In other words, they encapsulate the middle data between the extremes

The results for the larger airports (Deciles 6 and 9) coincide with one aspect of the initial theory that larger airports have more societal pressure to abide by and be environmentally-conscious of the airport emission levels. Some of the larger airports may have the funding in place to off-set the negative environmental impacts or better optimize processes to minimize their emissions internally. In other words, the increase in enplanements at larger capacity airports is correlated with an increase in good air quality. The results from Deciles 3 and 4 align with the other end of the theory that smaller airports face less public scrutiny, so their negative impact on air quality could be more significant than the impact of larger airports.

These observations are economically significant as enplanement capacity variations in counties have a different impact on air quality levels. This is reflected through the economic investment into the large-scale operations necessary for these airports that most likely include environmentally-conscious safeguards and plans to abide by local and state legislation. So, these larger airports anticipate an increase in enplanements and divert resources to remain below a threshold to avoid attracting negative publicity and attention.

Table 4 displays similar results to *Table 2* with the Quartiles of the enplanements, yet it includes control variables such as population and income. The number of observations may not be equal between the models as *Table 4* it uses the same categorization of the enplanement quartiles from *Table 2* to be consistent in the analysis.

Table 4: Two-Way-Fixed-Effects Approach for Quartiles 1 through 4 with Control

Regressors	Quartile 1	Quartile 2	Quartile 3	Quartile 4	All Counties
Ln (Enplanement)	0.00364 (0.01216)	-0.07221*** (0.01977)	0.18895*** (0.03414)	-0.01762 (0.02961)	-0.00849 (0.00762)
Ln (Population)	-0.50423*** (0.11171)	-0.26762*** (0.09955)	-0.92961*** (0.13288)	0.29326*** (0.10723)	-0.11175* (0.06006)
Ln (Income)	-0.47024*** (0.07878)	-0.24794*** (0.08294)	-0.08499 (0.10021)	-0.79460*** (0.09616)	-0.50376*** (0.04647)
Intercept	9.19442*** (1.17100)	6.38473*** (1.12837)	9.59636*** (1.83209)	3.75090** (1.55108)	6.24090*** (0.72483)
County Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Number of Observations	2,114	2,127	2,041	2,037	8,319
Adjusted R-Square	0.7334	0.8480	0.8778	0.9098	0.8882
Overall Significance	1.05E+10***	1.27E+02***	2.57E+09***	4.53E+02***	7.62E+09***

Source: EPA (2024), FAA (2024), with own calculations.
Notes: Robust standard Errors are in Parentheses. *, **, *** indicate 10%, 5%, and 1% significance levels, respectively and are clustered at county levels. The data is unbalanced as the number of counties in each state in a given year are not constant over time, as counties are added to the data or begin measuring AQI. Quartile 1=Model 1, etc. Observation numbers vary across the Quartiles because the same cut-off values are used from the enplanement data without control variables so the comparison is consistent between the tables. Control variables are Population and Income

As seen in *Table 4*, if enplanements in a county within Q2 go up by 1%, then the ratio of good air quality days goes down by about 0.07%. For Q3, if enplanements in a county increase by 1%, it is associated with an increase in the ratio of good air quality days by about 0.19%. These are slight changes from the coefficients in *Table 2* in values, yet the signs (positive and negative) and both having 1% statistical significance are the same between Quartiles 2 and 3 in

Table 2 and *Table 4*. This suggests that given the control variables, the airport size is a significant contributor to the increase or decrease in the ratio of good air quality days. However, the relationship for All Counties seemed to be insignificant after the control variables were added to the model.

Table 5 below demonstrates the results from *Table 3* with control variables such as Population and Income. Similar to the important note in the quartile tables, *Table 5* it uses the same categorization of the enplanement deciles from *Table 3* to be consistent in the analysis, so the observations may be different between the models. However, it seems as though the addition of the control variables when categorizing the enplanement data by decile, there is only one negative significant result at the 10% level for Decile 1. This could suggest that more control variables are needed.

Table 5: Log-Log Two-Way-Fixed-Effects Approach for 10 Deciles with Control

Regressors	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
Ln(Enplanement)	-0.04939*	0.03738	0.00785	-0.03870	-0.06171	-0.02365	0.03296	0.05019	-0.00288	-0.05672
	(0.02862)	(0.04508)	(0.03420)	(0.04414)	(0.04324)	(0.04079)	(0.03771)	(0.05670)	(0.03631)	(0.03812)
Ln(Population)	3.91798**	-1.51499*	0.65151	-0.75262**	0.98322	-0.18490	0.67915*	-1.14208**	1.66785***	3.98144**
	(1.53290)	(0.79582)	(0.79687)	(0.33703)	(0.74977)	(0.65776)	(0.39261)	(0.50558)	(0.49123)	(1.74411)
Ln(Income)	-0.33400	0.09541	0.07872	-0.47784*	0.55329**	-0.05128	0.81090**	-0.35673**	-0.28601	0.32378
	(0.27012)	(0.37865)	(0.30923)	(0.25656)	(0.25441)	(0.38115)	(0.34933)	(0.17623)	(0.22922)	(0.37779)
Intercept	-45.49598**	17.20578	-9.48828	14.65553***	-18.19544*	2.90140	-18.14568***	17.61767***	-18.88686***	-54.99046**
	(18.48321)	(11.34580)	(11.48583)	(5.06851)	(10.24502)	(9.05903)	(5.15532)	(6.38946)	(6.82082)	(23.27461)
County	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	864	864	864	864	863	864	864	864	864	544
Adjusted R-Square	0.9215	0.9837	0.9663	0.9314	0.9046	0.9669	0.9773	0.9788	0.9886	0.9946
Overall Significance	2.38E+07***	2.01E+07***	1.17E+05***	1.91E+08***	2.56E+05***	1.27E+08***	8.05E+05***	7.88E+05***	3.43E+07***	1.16E+08***

Source: EPA (2024), FAA (2024), with own calculations.

Notes: Robust standard Errors are in Parentheses. *, **, *** indicate 10%, 5%, and 1% significance levels, respectively and are clustered at county levels. The data is unbalanced as the number of counties in each state in a given year are not constant over time, as counties are added to the data or begin measuring AQI. Split into 10 Deciles between 10,000 to 53,515,982 enplanements; 1st Decile is Model 1, going up the decile ranges and 10th Decile is Model 10. Control Variables include Population, Income …

VIII. Conclusion

The purpose of this research was to determine the impact changes in airport capacity enplanements have on the respective air quality levels within the same county between 1999 and 2019. The main trends of the enplanement levels over the same period prove to be increasing. As for the type of air quality day, the ratio of good days increased and every other type of day decreased. So, analyzing the relationship between the two variables proves to be significant when evaluating the effect on air quality with different categorizations of airport capacity levels.

The increase in the large enplanement counties is correlated with an increase in percentage of good days throughout the year. In *Table 3*, Deciles 6 and 9 have positive statistically significant results of 0.125% and 0.262% respectively. However, the small-sized enplanement counties (Deciles 3 and 4) experienced a negative correlation between an increase in enplanement levels and a decrease in good air quality. As seen in *Table 2*, if enplanements in a county within Q2, between about 92,316 and 429,523 enplanements, go up by 1%, then the ratio of good air quality days goes down by 0.105% at a 1% significance level. For Q3, if enplanements in a county increase by 1%, it is associated with an increase in the ratio of good air quality days by 0.104% at a 1% significance level. This is also supported by the Deciles 3 and 4 results from *Table 3*, which demonstrates negative statistically significant outputs of 0.109% and 0.074% respectively. The results align with the initial theory of minimizing total cost and the societal pressures to abide by environmental laws for larger airports that may not apply to the smaller scale airports.

The implication of these results for policy makers is to enforce the environmental emission laws or amending them to ensure smaller-scale airports abide by them along with the larger airports. Another implication is providing more funding to smaller airports to assist in adjusting their operations to changes in enplanement to minimize environmental costs. In other words, continuing to enforce the environmental and air emission laws while expanding upon these laws to enforce them on a broader yet more specific scale.

Future researchers should continue this research post-COVID-19. During the pandemic shutdown, the airport capacity and enplanement levels dropped significantly. Once the quarantine period was lifted, the airport activity increased, and so did the airport emissions. It would be an interesting study to analyze the change in air quality once airports started returning to their pre-pandemic levels. A different angle would be comparing the enplanement capacity changes before and after pandemic and its effect on air quality.

IX. References

Ashok, A., Balakrishnan, H., & Barrett, S. R. H. (2017). Reducing the air quality and CO2 climate impacts of taxi and takeoff operations at airports. *Transportation Research Part D: Transport and Environment*, *54*, 287–303. https://doi.org/10.1016/j.trd.2017.05.013

Beketie, K. T., Angessa, A. T., Zeleke, T. T., & Ayal, D. Y. (2021). Impact of cement factory emission on air quality and human health around Mugher and the surrounding villages, Central Ethiopia. *Air Quality, Atmosphere & Health*, *15*(2), 347–361. https://doi.org/10.1007/s11869-021-01109-4

Edwards, H. A., Dixon-Hardy, D., & Wadud, Z. (2016). Aircraft cost index and the future of carbon emissions from Air Travel. *Applied Energy*, *164*, 553–562. https://doi.org/10.1016/j.apenergy.2015.11.058

Ekici, S., & Sevinc, H. (2021). Understanding a commercial airline company: A case study on emissions and air quality costs. *International Journal of Environmental Science and Technology*, *19*(6), 5139–5154. https://doi.org/10.1007/s13762-021-03471-3

Environmental Protection Agency. (n.d.). AirData website file download page. EPA. https://aqs.epa.gov/aqsweb/airdata/download_files.html Data extracted on February 19, 2024.

Hewings, G. J., Schindler, G. R., & Israilevich, P. R. (1997). Infrastructure and Economic Development: Airport Capacity in Chicago Region, 2001–18. *Journal of Infrastructure Systems*, *3*(3), 96–102. https://doi.org/10.1061/(ASCE)1076-0342(1997)3:3(96)

Hudda, N., Durant, L. W., Fruin, S. A., & Durant, J. L. (2020). Impacts of aviation emissions on near-airport residential air quality. *Environmental Science & Technology*, *54*(14), 8580–8588. https://doi.org/10.1021/acs.est.0c01859

Li, Q., Chen, W., Li, M., Yu, Q., & Wang, Y. (2022). Identifying the effects of industrial land expansion on PM2.5 concentrations: A Spatiotemporal Analysis in China. *Ecological Indicators*, *141*, 109069. https://doi.org/10.1016/j.ecolind.2022.109069

Luther, L. (2007, April 5). Environmental Impacts of Airport Operations, Maintenance, and Expansion. https://www.everycrsreport.com/files/20070405_RL33949_dcf1aeb939eebe185ccd8bf422201768a9fb6258.pdf

Nataraja, S., & Peterson, R. (2019). Direct economic impact analysis of the world's top five busiest airports in 2018. *International Journal of Business Administration*, *10*(6), 22. https://doi.org/10.5430/ijba.v10n6p22

Passenger boarding (enplanement) and all-cargo data for U.S. airports - previous years. Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports - Previous Years | Federal Aviation Administration. (n.d.). https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/previous_years#2000 Data extracted on February 19, 2024.

Total population: All ages including Armed Forces Overseas 1999-2019. FRED. (2023, December 19). https://fred.stlouisfed.org/series/POP#0 Data extracted on March 9, 2024.

United States Cities Database. simplemaps. (n.d.). https://simplemaps.com/data/us-cities Data extracted on February 19, 2024.

Wolfe, P. J., Yim, S. H. L., Lee, G., Ashok, A., Barrett, S. R. H., & Waitz, I. A. (2014). Near-airport distribution of the environmental costs of Aviation. *Transport Policy*, *34*, 102–108. https://doi.org/10.1016/j.tranpol.2014.02.023

Zheng, X., Peng, W., & Hu, M. (2020). Airport noise and house prices: A quasi-experimental design study. *Land Use Policy*, *90*, 104287. https://doi.org/10.1016/j.landusepol.2019.104287

Zhu, Y., Fanning, E., Yu, R. C., Zhang, Q., & Froines, J. R. (2011). Aircraft emissions and local air quality impacts from takeoff activities at a large International Airport. *Atmospheric Environment*, *45*(36), 6526–6533. https://doi.org/10.1016/j.atmosenv.2011.08.062

X. Appendix

```
libname AEData "~/my_shared_file_links/u47408605/Data"
access=readonly;
run;

proc import datafile="/home/u63024456/MySAS/AQI Data Analysis/ALL COUNTY DATA
COMPILED.xlsx"
	out=work.ECON1
	dbms=xlsx
	replace;
	sheet="ALL COUNTY AQI COMPILED";
	getnames=yes;
run;

proc sort data=ECON1;
	by County_State year;
Run;

/*Summary Statistics*/
proc means data=ECON1;
	 var GD MD USGD UD VUD HD;
run;

/* proc means data=ECON;
	 var GD MD USGD UD VUD HD;
run; */

/*Maybe
proc means data=ECON;
	 var GD MD USGD UD VUD HD;
	 by County_State;
run;*/



/* Enplanement Data*/
proc import datafile="/home/u63024456/MySAS/AQI Data Analysis/Enplanement Data
Compiled Updated Again.xlsx"
	out=work.PLANE3
	dbms=xlsx
	replace;
	getnames=yes;
	sheet="All Enplanement Data";
run;
```

```
/* Sort Enplanement to the aggregate*/
proc sort data=PLANE3;
        by County_State year;
Run;

/* PLANE 4 has 23,618 observations */
PROC SQL;
create table PLANE4 as
select County_State, year, LOCID,
        sum(CY_Enplanement) as Agg_CY_Enplanement,
        sum(PY_Enplanement) as Agg_PY_Enplanement
from PLANE3
group by County_State, year;
QUIT;

/* Find the highest enplanement county and year */
proc sort data=PLANE4;
        by descending Agg_CY_Enplanement;
run;

Proc Print Data=PLANE3 (obs=10);
run;

Proc Print Data=PLANE4 (obs=8);
run;

/* Initial before removing NA counties*/
/* Summary Statistics for Enplanement */
proc sort data=PLANE4;
        by year;
run;

proc means data=PLANE4;
         var Agg_CY_Enplanement;
         *by Year;
run;



/* Observations for PLANE 5 is 13,025 */
data PLANE5;
set PLANE4;
where (LOCID ne "GUM") and (LOCID ne "UAM") and (LOCID ne "6Y8")
        and (LOCID ne "AWK") and (LOCID ne "CPX") and (LOCID ne "FAQ")
        and (LOCID ne "GRO") and (LOCID ne "GSN") and (LOCID ne "HTO")
        and (LOCID ne "JON") and (LOCID ne "KWF") and (LOCID ne "NRR")
```

```
        and (LOCID ne "NTD") and (LOCID ne "ORS") and (LOCID ne "PPC")
        and (LOCID ne "PPG") and (LOCID ne "SJX") and (LOCID ne "STT")
        and (LOCID ne "STX") and (LOCID ne "SXP") and (LOCID ne "TNI")
        and (LOCID ne "UAM") and (LOCID ne "VI22") and (LOCID ne "VI32")
        and (LOCID ne "X67") and (LOCID ne "PPG") and (LOCID ne "NSI")
        and (LOCID ne "LBX") and (LOCID ne "GSN") and (LOCID ne "HTO")
        and (County_State ne "#N/A")
        and (Agg_CY_Enplanement>10000);
run;

proc sort data=PLANE5;
        by descending Agg_CY_Enplanement;
run;

proc sort data=PLANE5;
        by County_State;
run;


/* New proc means for PLANE w/o missing values */
proc means data=PLANE5;
         var Agg_CY_Enplanement;

         *by Year;
run;


/*  QUARTILES  */
proc univariate data=PLANE5;
   var Agg_CY_Enplanement;
   output out=quartile_data
   pctlpts = 25 50 75
   pctlpre = Q_;
run;


/* Get rid of missing values */
data ECON2;
set ECON1;
where (County_State ne "#N/A");
run;

proc sort data=ECON2;
        by County_State;
run;
```

```
/* MERGE DATABASES */
proc sort data=ECON2;
	by Year County_State;
run;

proc sort data=PLANE5;
	by Year County_State;
run;

/* Observations 29,556  */
data Combined;
	merge ECON2 PLANE5;
	by Year County_State;
	*if FirstEffectiveYear="-" then DID=0;
		*else if Year>=FirstEffectiveYear then DID=1;
		*else DID=0;
	lnEnplanement = log(Agg_CY_Enplanement+1);
	lnGD = log(GD);
	keep County_State Year lnEnplanement lnGD Agg_CY_Enplanement
Agg_PY_Enplanement GD MD USGD UD VUD HD;
run;


/* Separate data by 10% for observations */
proc sort data=Combined;
	by Agg_CY_Enplanement;
run;

data Combined2;
set Combined;
where (Agg_CY_Enplanement ne .) and (GD ne .);
*where GD ne .;
	id=_N_;
run;


proc sort data=Combined;
	by descending Agg_CY_Enplanement;
run;


/* PROC SURVEYREG PROCEDURES FOR 6 AQI */
/* Good Days */
ods output ParameterEstimates=PEforGD DataSummary=ObsGD
```

```
                              FitStatistics=AdjRsqGD Effects=OverallSigGD;
Proc SurveyReg Data=Combined;
          Class County_State Year / ref=First;
          Model GD = Agg_CY_Enplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;

/* Moderate Days */
ods output ParameterEstimates=PEforMD DataSummary=ObsMD
                              FitStatistics=AdjRsqMD Effects=OverallSigMD;
Proc SurveyReg Data=Combined;
          Class County_State Year / ref=First;
          Model MD = Agg_CY_Enplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;

/* USGD Days */
ods output ParameterEstimates=PEforUSGD DataSummary=ObsUSGD
                              FitStatistics=AdjRsqUSGD Effects=OverallSigUSGD;
Proc SurveyReg Data=Combined;
          Class County_State Year / ref=First;
          Model USGD = Agg_CY_Enplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;

/* Unhealthy Days */
ods output ParameterEstimates=PEforUD DataSummary=ObsUD
                              FitStatistics=AdjRsqUD Effects=OverallSigUD;
Proc SurveyReg Data=Combined;
          Class County_State Year / ref=First;
          Model UD = Agg_CY_Enplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;

/* Very Unhealthy Days */
ods output ParameterEstimates=PEforVUD DataSummary=ObsVUD
                              FitStatistics=AdjRsqVUD Effects=OverallSigVUD;
Proc SurveyReg Data=Combined;
          Class County_State Year / ref=First;
          Model VUD = Agg_CY_Enplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;

/* Hazardous Days */
ods output ParameterEstimates=PEforHD DataSummary=ObsHD
                              FitStatistics=AdjRsqHD Effects=OverallSigHD;
```

```
Proc SurveyReg Data=Combined;
	Class County_State Year / ref=First;
	Model HD = Agg_CY_Enplanement County_State Year / Solution AdjRsq;
	*cluster County_State;
run;


/************************************************************
************** MODELS CUTTING ENPLANEMENT DATA ***************
************************************************************/

proc means data=Combined;
	var Agg_CY_Enplanement;
run;


proc means data=Combined2 P25 P50 P75;
	var id Agg_CY_Enplanement;
run;


/*********************************************************

                    Quartile Percentile Results

**********************************************************/
ods output ParameterEstimates=PEforQ4 DataSummary=ObsQ4
			FitStatistics=AdjRsqQ4 Effects=OverallSigQ4;
Proc SurveyReg Data=Combined2;
	where (Agg_CY_Enplanement>3416745);
	*where (id > 6479);
	Class County_State Year / ref=First;
	ModelA: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
	*cluster County_State;
run;


ods output ParameterEstimates=PEforQ3 DataSummary=ObsQ3
			FitStatistics=AdjRsqQ3 Effects=OverallSigQ3;
Proc SurveyReg Data=Combined2;
	where (3416745>=Agg_CY_Enplanement>429523);
	*where (6479>=id>4319.5);
	Class County_State Year / ref=First;
	ModelB: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
	*cluster County_State;
run;
```

```
ods output ParameterEstimates=PEforQ2 DataSummary=ObsQ2
                              FitStatistics=AdjRsqQ2 Effects=OverallSigQ2;
Proc SurveyReg Data=Combined2;
          where (429523>=Agg_CY_Enplanement>92316);
          *where (4319.5>=id>2160);
          Class County_State Year / ref=First;
          ModelC: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;


ods output ParameterEstimates=PEforQ1 DataSummary=ObsQ1
                              FitStatistics=AdjRsqQ1 Effects=OverallSigQ1;
Proc SurveyReg Data=Combined2;
          where (92316>=Agg_CY_Enplanement);
          *where (2160>=id);
          Class County_State Year / ref=First;
          ModelD: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;


ods output ParameterEstimates=PEforAll DataSummary=ObsAll
                              FitStatistics=AdjRsqAll Effects=OverallSigAll;
Proc SurveyReg Data=Combined2;
          Class County_State Year / ref=First;
          ModelE: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;



Data NewData;
          Set Combined2;
          *Where (Agg_CY_Enplanement>1233883);
          Where (1233883>=Agg_CY_Enplanement>196230);
          *where (196230>=Agg_CY_Enplanement>51639);
          *where (51639>=Agg_CY_Enplanement);


          /*This should be one of the where statements you are using in proc surveyreg. */
Run;
```

```
/****************************************

RESULTS FOR QUARTILES

**********************************/

Data Table_Long2;
	length Model $10;
	length Parameter $30;
	set PEforQ1 PEforQ2 PEforQ3 PEforQ4 PEforAll indsname=M;
	*THisISM=M;
	where Parameter="Intercept" or Parameter="lnEnplanement"/*or Parameter="DID" */;

	if	M="WORK.PEFORQ1" then Model="Model1";
		else if M="WORK.PEFORQ2" then Model="Model2";
		else if M="WORK.PEFORQ3" then Model="Model3";
		else if M="WORK.PEFORQ4" then Model="Model4";
		else if M="WORK.PEFORALL" then Model="Model5";

	if Probt le 0.01 then Star="***";
		else if Probt le 0.05 then Star="**";
		else if Probt le 0.1 then Star="*";

	EditedResults=cats(Put(Estimate,comma16.5),star);
	output;

	EditedResults=cats("(",put(StdErr,comma16.5),")");
	output;

run;



proc sort data=Table_Long2 out=Table_Long_Sorted2;
	by Model Parameter;
run;


/* Step 2: Create separate results columns for each model */
data Model1Results(rename=(EditedResults=Model1))
	Model2Results(rename=(EditedResults=Model2))
	Model3Results(rename=(EditedResults=Model3))
	Model4Results(rename=(EditedResults=Model4))
```

```
        Model5Results(rename=(EditedResults=Model5));
        set Table_Long_Sorted2;

        if Model="Model1" then output Model1Results;
                else if Model="Model2" then output Model2Results;
                else if Model="Model3" then output Model3Results;
                else if Model="Model4" then output Model4Results;
                else if Model="Model5" then output Model5Results;
        drop Model;
        keep Parameter EditedResults;
run;

data Table_Wide2;
        merge Model1Results Model2Results Model3Results Model4Results Model5Results;
        by Parameter;

        if Parameter="lnEnplanement" then Order=1;
                else if substr(Parameter,1,9)="Intercept" then Order=2;

        if mod(_n_,2)=1 then Regressors=Parameter;

run;


/* Order the variables in the results table */
proc sort data=Table_Wide2 out=Table_Wide_Sorted2(drop=Order Parameter);
        by Order;
run;


/* Step 4: Create the rows for other statistics */
/* County and Year Fixed Effects */
Data Control2;
        Regressors="County";
        Model1="Yes";
        Model2="Yes";
        Model3="Yes";
        Model4="Yes";
        Model5="Yes";

        output;
        Regressors="Year";
        output;

run;
```

```
/* The row for the number of observations */
Data NumofObs2;
        merge ObsQ1(rename=(Nvalue1=NVModel1) drop=CValue1)
                        ObsQ2(rename=(Nvalue1=NVModel2) drop=CValue1)
                        ObsQ3(rename=(Nvalue1=NVModel3) drop=CValue1)
                        ObsQ4(rename=(Nvalue1=NVModel4) drop=CValue1)
                        ObsAll(rename=(Nvalue1=NVModel5) drop=CValue1);
        where Label1="Number of Observations";
        Model1=put(NVModel1,comma16.);
        Model2=put(NVModel2,comma16.);
        Model3=put(NVModel3,comma16.);
        Model4=Put(NVModel4,comma16.);
        Model5=Put(NVModel5,comma16.);

        keep Label1 Model1 Model2 Model3 Model4 Model5;
run;

/* The row for the adjusted R-Squared */
Data AdjRsq2;
        merge AdjRsqQ1(rename=(cvalue1=Model1) drop=nvalue1)
                AdjRsqQ2(rename=(cvalue1=Model2) drop=nvalue1)
                AdjRsqQ3(rename=(cvalue1=Model3) drop=nvalue1)
                AdjRsqQ4(rename=(cvalue1=Model4) drop=nvalue1)
                AdjRsqAll(rename=(cvalue1=Model5) drop=nvalue1);
        Where Label1="Adjusted R-Square";
run;

/* The row for the F-test related to the Overall Significance of the model */
Data OSM1(rename=(EditedValue=Model1)) OSM2(rename=(EditedValue=Model2))
OSM3(rename=(EditedValue=Model3))
                OSM4(rename=(EditedValue=Model4)) OSM5(rename=(EditedValue=Model5));
        set OverallSigQ1 OverallSigQ2 OverallSigQ3 OverallSigQ4 OverallSigAll
indsname=M;
        where Effect="Model";
        if ProbF le 0.01 then Star="***";
                else if ProbF le 0.05 then Star="**";
                else if ProbF le 0.1 then Star="*";
        ThisIsM=M;

        Label1="Overall Significance";
        *EditedValue=cats(put(FValue,BESTw.2),Star);
        EditedValue=cats(put(FValue,e9.),Star);

                if M="WORK.OVERALLSIGQ1" then output OSM1;
                else if M="WORK.OVERALLSIGQ2" then output OSM2;
```

```
            else if M="WORK.OVERALLSIGQ3" then output OSM3;
            else if M="WORK.OVERALLSIGQ4" then output OSM4;
            else if M="WORK.OVERALLSIGALL" then output OSM5;

    keep Label1 EditedValue;
run;


Data OverallSig2;
    merge OSM1 OSM2 OSM3 OSM4 OSM5;
    by Label1;
run;



/* Combine all rows for other statistics */
data OtherStat2;
    set NumofObs2 AdjRsq2 OverallSig2;
    rename Label1=Regressors;
run;

/* Add rows for other statistics to the table */
Data Table_Wide_Sorted_withStat3;
    set Table_Wide_Sorted2 Control2 OtherStat2;
run;



/* Print the clean results table */
/* New Code: The name of the excel file, the title of the results table, and its footnote are
modified */
ods excel file="/home/u63024456/MySAS/AQI Data
Analysis/FinalTableGDEnplanementQUARTILES.xlsx" options(Embedded_Titles="ON"
Embedded_Footnotes="ON"); /*Use the path to your MySAS folder */
Title "Table 2: Two-Way-Fixed-Effects Approach for Quartiles 1 through 4";
footnote justify=left "Source: EPA (2024), FAA (2024), with own calculations.";
footnote2 justify=left "Notes: Robust standard Errors are in Parentheses. *, **, *** indicate 10%,
5%,
    and 1% significance levels, respectively and are clustered at county levels. The data is
    unbalanced as the number of counties in each state in a given year are not constant over
time,
    as counties are added to the data or begin measuring AQI. Quartile 1=Model 1, etc.";
proc print data=Table_Wide_Sorted_withStat3 noobs;
    var Regressors;
```

```
	var Model1 Model2 Model3 Model4 Model5/ style(header)={Just=Center}
style(data)={Just=Center tagattr="type:string"};

	format Regressors $VariableName.;
run;
ods excel close;



/*******************************************************

				10*10 Percentile Results

*******************************************************/


proc means data=Combined2 P10 P20 P30 P40 P50 P60 P70 P80 P90;
	var id Agg_CY_Enplanement;
run;


/* (A 90th percentile) enplanements on GD */
ods output ParameterEstimates=PEforGDA DataSummary=ObsGDA
			FitStatistics=AdjRsqGDA Effects=OverallSigGDA;
Proc SurveyReg Data=Combined2;
	*where (Agg_CY_Enplanement>17274506);
	where (id>7775);
	Class County_State Year / ref=First;
	ModelA: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
	*cluster County_State;
run;

/* (B 80th percentile) enplanement on GD */
ods output ParameterEstimates=PEforGDB DataSummary=ObsGDB
			FitStatistics=AdjRsqGDB Effects=OverallSigGDB;
Proc SurveyReg Data=Combined2;
	*where (17274506>=Agg_CY_Enplanement>5054191);
	where (7775>=id>6911);
	Class County_State Year / ref=First;
	ModelB: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
	*cluster County_State;
run;


/* (C 70th percentile) Enplanement on GD */
```

```
ods output ParameterEstimates=PEforGDC DataSummary=ObsGDC
                              FitStatistics=AdjRsqGDC Effects=OverallSigGDC;
Proc SurveyReg Data=Combined2;
          *where (5054191>=Agg_CY_Enplanement>2128405);
          where (6911>=id>6047);
          Class County_State Year / ref=First;
          ModelC: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;

/* (D 60th percentile) Enplanement on GD*/
ods output ParameterEstimates=PEforGDD DataSummary=ObsGDD
                              FitStatistics=AdjRsqGDD Effects=OverallSigGDD;
Proc SurveyReg Data=Combined2;
          *where (2128405>=Agg_CY_Enplanement>846369);
          where (6047>=id>5183);
          Class County_State Year / ref=First;
          ModelD: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;

/* (E 50th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDE DataSummary=ObsGDE
                              FitStatistics=AdjRsqGDE Effects=OverallSigGDE;
Proc SurveyReg Data=Combined2;
          *where (846369>=Agg_CY_Enplanement>429523);
          where (5183>=id>4319.5);
          Class County_State Year / ref=First;
          ModelE: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;


/* (F 40th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDF DataSummary=ObsGDF
                              FitStatistics=AdjRsqGDF Effects=OverallSigGDF;
Proc SurveyReg Data=Combined2;
          *where (429523>=Agg_CY_Enplanement>263217);
          where (4319.5>=id>3456);
          Class County_State Year / ref=First;
          ModelF: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
          *cluster County_State;
run;


/* (G 30th percentile) Enplanement on GD */
```

```
ods output ParameterEstimates=PEforGDG DataSummary=ObsGDG
                        FitStatistics=AdjRsqGDG Effects=OverallSigGDG;
Proc SurveyReg Data=Combined2;
        *where (263217>=Agg_CY_Enplanement>138423);
        where (3456>=id>2592);
        Class County_State Year / ref=First;
        ModelG: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
        *cluster County_State;
run;

/* (H 20th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDH DataSummary=ObsGDH
                        FitStatistics=AdjRsqGDH Effects=OverallSigGDH;
Proc SurveyReg Data=Combined2;
        *where (138423>=Agg_CY_Enplanement>54728);
        where (2592>=id>1728);
        Class County_State Year / ref=First;
        ModelH: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
        *cluster County_State;
run;

/* (I 20th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDI DataSummary=ObsGDI
                        FitStatistics=AdjRsqGDI Effects=OverallSigGDI;
Proc SurveyReg Data=Combined2;
        *where (54728>=Agg_CY_Enplanement>23240);
        where (1728>=id>864);
        Class County_State Year / ref=First;
        ModelI: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
        *cluster County_State;
run;

/* (J 10th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDJ DataSummary=ObsGDJ
                        FitStatistics=AdjRsqGDJ Effects=OverallSigGDJ;
Proc SurveyReg Data=Combined2;
        *where (23240>=Agg_CY_Enplanement);
        where (864>=id);
        Class County_State Year / ref=First;
        ModelJ: Model lnGD = lnEnplanement County_State Year / Solution AdjRsq;
        *cluster County_State;
run;
```

```
/* BUILD RESULTS */

Data Table_Long;
	length Model $10;
	length Parameter $30;
	set PEforGDA PEforGDB PEforGDC PEforGDD PEforGDE PEforGDF PEforGDG
PEforGDH PEforGDI PEforGDJ indsname=M;
	*THisISM=M;
	where Parameter="Intercept" or Parameter="lnEnplanement"/*or Parameter="DID" */;

	if	M="WORK.PEFORGDA" then Model="Model10";
		else if M="WORK.PEFORGDB" then Model="Model9";
		else if M="WORK.PEFORGDC" then Model="Model8";
		else if M="WORK.PEFORGDD" then Model="Model7";
		else if M="WORK.PEFORGDE" then Model="Model6";
		else if M="WORK.PEFORGDF" then Model="Model5";
		else if M="WORK.PEFORGDG" then Model="Model4";
		else if M="WORK.PEFORGDH" then Model="Model3";
		else if M="WORK.PEFORGDI" then Model="Model2";
		else if M="WORK.PEFORGDJ" then Model="Model1";

	if Probt le 0.01 then Star="***";
		else if Probt le 0.05 then Star="**";
		else if Probt le 0.1 then Star="*";

	EditedResults=cats(Put(Estimate,comma16.5),star);
	output;

	EditedResults=cats("(",put(StdErr,comma16.5),")");
	output;

run;



proc sort data=Table_Long out=Table_Long_Sorted;
	by Model Parameter;
run;


/* Step 2: Create separate results columns for each model */
data Model1Results(rename=(EditedResults=Model1))
	Model2Results(rename=(EditedResults=Model2))
	Model3Results(rename=(EditedResults=Model3))
```

```
	Model4Results(rename=(EditedResults=Model4))
	Model5Results(rename=(EditedResults=Model5))
	Model6Results(rename=(EditedResults=Model6))
	Model7Results(rename=(EditedResults=Model7))
	Model8Results(rename=(EditedResults=Model8))
	Model9Results(rename=(EditedResults=Model9))
	Model10Results(rename=(EditedResults=Model10));
	set Table_Long_Sorted;

	if Model="Model1" then output Model1Results;
		else if Model="Model2" then output Model2Results;
		else if Model="Model3" then output Model3Results;
		else if Model="Model4" then output Model4Results;
		else if Model="Model5" then output Model5Results;
		else if Model="Model6" then output Model6Results;
		else if Model="Model7" then output Model7Results;
		else if Model="Model8" then output Model8Results;
		else if Model="Model9" then output Model9Results;
		else if Model="Model10" then output Model10Results;
	drop Model;
	keep Parameter EditedResults;
run;

data Table_Wide;
	merge Model1Results Model2Results Model3Results Model4Results Model5Results
Model6Results
	Model7Results Model8Results Model9Results Model10Results;
	by Parameter;

	if Parameter="lnEnplanement" then Order=1;
		else if substr(Parameter,1,9)="Intercept" then Order=2;

	*if Parameter="DID" then Order=1;
		*else if substr(Parameter,1,9)="Intercept" then Order=2;

	if mod(_n_,2)=1 then Regressors=Parameter;

run;


/* Order the variables in the results table */
proc sort data=Table_Wide out=Table_Wide_Sorted(drop=Order Parameter);
	by Order;
run;
```

```
/* Step 4: Create the rows for other statistics */
/* County and Year Fixed Effects */
Data Control;
	Regressors="County";
	Model1="Yes";
	Model2="Yes";
	Model3="Yes";
	Model4="Yes";
	Model5="Yes";
	Model6="Yes";
	Model7="Yes";
	Model8="Yes";
	Model9="Yes";
	Model10="Yes";
	output;
	Regressors="Year";
	output;

run;


/* The row for the number of observations */
Data NumofObs;
	merge ObsGDA(rename=(Nvalue1=NVModel10) drop=CValue1)
			ObsGDB(rename=(Nvalue1=NVModel9) drop=CValue1)
			ObsGDC(rename=(Nvalue1=NVModel8) drop=CValue1)
			ObsGDD(rename=(Nvalue1=NVModel7) drop=CValue1)
			ObsGDE(rename=(Nvalue1=NVModel6) drop=CValue1)
			ObsGDF(rename=(Nvalue1=NVModel5) drop=CValue1)
			ObsGDG(rename=(Nvalue1=NVModel4) drop=CValue1)
			ObsGDH(rename=(Nvalue1=NVModel3) drop=CValue1)
			ObsGDI(rename=(Nvalue1=NVModel2) drop=CValue1)
			ObsGDJ(rename=(Nvalue1=NVModel1) drop=CValue1);
	where Label1="Number of Observations";
	Model1=put(NVModel1,comma16.);
	Model2=put(NVModel2,comma16.);
	Model3=put(NVModel3,comma16.);
	Model4=Put(NVModel4,comma16.);
	Model5=put(NVModel5,comma16.);
	Model6=put(NVModel6,comma16.);
	Model7=put(NVModel7,comma16.);
	Model8=put(NVModel8,comma16.);
	Model9=put(NVModel9,comma16.);
	Model10=put(NVModel10,comma16.);
	keep Label1 Model1 Model2 Model3 Model4 Model5 Model6 Model7 Model8 Model9
Model10;
```

```
run;

/* The row for the adjusted R-Squared */
Data AdjRsq;
	merge AdjRsqGDA(rename=(cvalue1=Model10) drop=nvalue1)
		AdjRsqGDB(rename=(cvalue1=Model9) drop=nvalue1)
		AdjRsqGDC(rename=(cvalue1=Model8) drop=nvalue1)
		AdjRsqGDD(rename=(cvalue1=Model7) drop=nvalue1)
		AdjRsqGDE(rename=(cvalue1=Model6) drop=nvalue1)
		AdjRsqGDF(rename=(cvalue1=Model5) drop=nvalue1)
		AdjRsqGDG(rename=(cvalue1=Model4) drop=nvalue1)
		AdjRsqGDH(rename=(cvalue1=Model3) drop=nvalue1)
		AdjRsqGDI(rename=(cvalue1=Model2) drop=nvalue1)
		AdjRsqGDJ(rename=(cvalue1=Model1) drop=nvalue1);
	Where Label1="Adjusted R-Square";
run;

/* The row for the F-test related to the Overall Significance of the model */
Data OSM1(rename=(EditedValue=Model1)) OSM2(rename=(EditedValue=Model2))
OSM3(rename=(EditedValue=Model3))
		OSM4(rename=(EditedValue=Model4)) OSM5(rename=(EditedValue=Model5))
OSM6(rename=(EditedValue=Model6))
		OSM7(rename=(EditedValue=Model7)) OSM8(rename=(EditedValue=Model8))
OSM9(rename=(EditedValue=Model9))
		OSM10(rename=(EditedValue=Model10));
	set OverallSigGDA OverallSigGDB OverallSigGDC OverallSigGDD OverallSigGDE
OverallSigGDF
	OverallSigGDG OverallSigGDH OverallSigGDI OverallSigGDJ indsname=M;
	where Effect="Model";
	if ProbF le 0.01 then Star="***";
		else if ProbF le 0.05 then Star="**";
		else if ProbF le 0.1 then Star="*";
	ThisIsM=M;

	Label1="Overall Significance";
	*EditedValue=cats(put(FValue,BESTw.2),Star);
	EditedValue=cats(put(FValue,e9.),Star);

		if M="WORK.OVERALLSIGGDA" then output OSM10;
		else if M="WORK.OVERALLSIGGDB" then output OSM9;
		else if M="WORK.OVERALLSIGGDC" then output OSM8;
		else if M="WORK.OVERALLSIGGDD" then output OSM7;
		else if M="WORK.OVERALLSIGGDE" then output OSM6;
		else if M="WORK.OVERALLSIGGDF" then output OSM5;
		else if M="WORK.OVERALLSIGGDG" then output OSM4;
		else if M="WORK.OVERALLSIGGDH" then output OSM3;
```

```
		else if M="WORK.OVERALLSIGGDI" then output OSM2;
		else if M="WORK.OVERALLSIGGDJ" then output OSM1;
	keep Label1 EditedValue;
run;


Data OverallSig;
	merge OSM1 OSM2 OSM3 OSM4 OSM5 OSM6 OSM7 OSM8 OSM9 OSM10;
	by Label1;
run;



/* Combine all rows for other statistics */
data OtherStat;
	set NumofObs AdjRsq OverallSig;
	rename Label1=Regressors;
run;

/* Add rows for other statistics to the table */
Data Table_Wide_Sorted_withStat;
	set Table_Wide_Sorted Control OtherStat;
run;



/* Print the clean results table */
/* New Code: The name of the excel file, the title of the results table, and its footnote are
modified */
ods excel file="/home/u63024456/MySAS/AQI Data
Analysis/FinalTableGDEnplanementPERCENTILES.xlsx" options(Embedded_Titles="ON"
Embedded_Footnotes="ON"); /*Use the path to your MySAS folder */
Title "Table 3: Two-Way-Fixed-Effects Approach for 10 Deciles";
footnote justify=left "Source: EPA (2024), FAA (2024), with own calculations.";
footnote2 justify=left "Robust standard Errors are in Parentheses. *, **, *** indicate 10%, 5%,
and 1% significance
				levels, respectively and are clustered at county levels. The data is
				unbalanced as the number of counties in each state in a given year
are not constant over time,
				as counties are added to the data or begin measuring AQI. 1st
Decile is represented by Model 1,
				going up the decile ranges as the Model number increases to the
10th Decile for Model 10.";
proc print data=Table_Wide_Sorted_withStat noobs;
	var Regressors;
```

```
	var Model1 Model2 Model3 Model4 Model5 Model6 Model7 Model8 Model9 Model10 / 
style(header)={Just=Center} style(data)={Just=Center tagattr="type:string"};

	format Regressors $VariableName.;
run;
ods excel close;




/* Both Tables  */

ods excel file="/home/u63024456/MySAS/AQI Data 
Analysis/FinalTableGDEnplanementQUARTILES.xlsx" options(Embedded_Titles="ON" 
Embedded_Footnotes="ON"); /*Use the path to your MySAS folder */
Title "Table 2: Two-Way-Fixed-Effects Approach for Quartiles 1 through 4";
footnote justify=left "Source: EPA (2024), FAA (2024), with own calculations.";
footnote2 justify=left "Notes: Robust standard Errors are in Parentheses. *, **, *** indicate 10%, 
5%,
	and 1% significance levels, respectively and are clustered at county levels. The data is 
	unbalanced as the number of counties in each state in a given year are not constant over 
time,
	as counties are added to the data or begin measuring AQI. Quartile 1=Model 1, etc.";
proc print data=Table_Wide_Sorted_withStat3 noobs;
	var Regressors;

	var Model1 Model2 Model3 Model4 / style(header)={Just=Center} 
style(data)={Just=Center tagattr="type:string"};

	format Regressors $VariableName.;
run;
ods excel close;




/* Print the clean results table */
/* New Code: The name of the excel file, the title of the results table, and its footnote are 
modified */
ods excel file="/home/u63024456/MySAS/AQI Data 
Analysis/FinalTableGDEnplanementPERCENTILES.xlsx" options(Embedded_Titles="ON" 
Embedded_Footnotes="ON"); /*Use the path to your MySAS folder */
Title "Table 3: Two-Way-Fixed-Effects Approach for 10 Percentiles";
footnote justify=left "Source: EPA (2024), FAA (2024), with own calculations.";
footnote2 justify=left "Notes: Robust standard Errors are in Parentheses.
```

```
                              *, **, *** indicate 10%, 5%, and 1% significance levels, respectively and
are clustered at county levels.
                              The data is unbalanced as the number of counties in each state in a given
year are not constant over time,
                              as counties are added to the data or begin measuring AQI.";
proc print data=Table_Wide_Sorted_withStat noobs;
	var Regressors;

	var Model1 Model2 Model3 Model4 Model5 Model6 Model7 Model8 Model9 Model10 /
style(header)={Just=Center} style(data)={Just=Center tagattr="type:string"};

	format Regressors $VariableName.;
run;
ods excel close;




/*
*******************************************************
*******************************************************
*******************************************************

                    ADD CONTROL VARIABLES

*******************************************************
*******************************************************
*******************************************************
*/


/* Income and Population Controls */
proc import datafile="/home/u63024456/MySAS/AQI Data Analysis/Income_Control.xlsx"
	out=work.Income
	dbms=xlsx
	replace;
	getnames=yes;
	sheet="Sheet1";
run;

proc sort data=Income;
	by County_State Year;
Run;
```

```
/*Summary Statistics*/
proc means data=Income;
           var Income;
           *var Population;
run;



/* vehicle mileage, gasoline prices, precipitation  */








/*************************************************************

                    COMBINE ALL DATABASES WITH CONTROL VARIABLES


*************************************************************
*************************************************************/
proc sort data=ECON2;
          by Year County_State;
run;

proc sort data=PLANE5;
          by Year County_State;
run;

proc sort data=Income;
          by Year County_State;
run;



/* ALL THE DATA BELOW  */
data CombinedControl;
          merge ECON2 PLANE5 Income;
          by Year County_State;
          lnEnplanement = log(Agg_CY_Enplanement+1);
          lnGD = log(GD);
          lnIncome = log(Income);
```

```
        lnPopulation = log(Population);

        keep County_State Year lnEnplanement lnGD Agg_CY_Enplanement
Agg_PY_Enplanement GD MD USGD UD VUD HD
        lnIncome lnPopulation;
run;

proc sort Data=CombinedControl;
by lnIncome;
run;

data CombinedControlFinal;
set CombinedControl;
where (Agg_CY_Enplanement ne .) and (GD ne .) and (lnIncome ne .) and (lnPopulation ne .);
*where GD ne .;
        id=_N_;
run;

proc means data=CombinedControlFinal P25 P50 P75;
        var id Agg_CY_Enplanement;
run;


/***************************************************************
****************************************************************
                    MODELS 1-4 FOR QUARTILES
****************************************************************
****************************************************************/
ods output ParameterEstimates=PEforQ4CC DataSummary=ObsQ4CC
                FitStatistics=AdjRsqQ4CC Effects=OverallSigQ4CC;
Proc SurveyReg Data=CombinedControlFinal;
        where (Agg_CY_Enplanement>3416745);
        *where (id > 6479);
        Class County_State Year / ref=First;
        ModelA: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
run;


ods output ParameterEstimates=PEforQ3CC DataSummary=ObsQ3CC
                FitStatistics=AdjRsqQ3CC Effects=OverallSigQ3CC;
Proc SurveyReg Data=CombinedControlFinal;
        where (3416745>=Agg_CY_Enplanement>429523);
        *where (6479>=id>4319.5);
        Class County_State Year / ref=First;
```

```
	ModelB: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
	*cluster County_State;
run;


ods output ParameterEstimates=PEforQ2CC DataSummary=ObsQ2CC
			FitStatistics=AdjRsqQ2CC Effects=OverallSigQ2CC;
Proc SurveyReg Data=CombinedControlFinal;
	where (429523>=Agg_CY_Enplanement>92316);
	*where (4319.5>=id>2160);
	Class County_State Year / ref=First;
	ModelC: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
	*cluster County_State;
run;


ods output ParameterEstimates=PEforQ1CC DataSummary=ObsQ1CC
			FitStatistics=AdjRsqQ1CC Effects=OverallSigQ1CC;
Proc SurveyReg Data=CombinedControlFinal;
	where (92316>=Agg_CY_Enplanement);
	*where (2160>=id);
	Class County_State Year / ref=First;
	ModelD: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
	*cluster County_State;
run;

ods output ParameterEstimates=PEforAllCC DataSummary=ObsAllCC
			FitStatistics=AdjRsqAllCC Effects=OverallSigAllCC;
Proc SurveyReg Data=CombinedControlFinal;
	Class County_State Year / ref=First;
	ModelE: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
	*cluster County_State;
run;




/*****************************************************************
******************************************************************
				RESULTS FOR QUARTILES
******************************************************************
```

```
*****************************************************************/

Data Table_LongCC;
	length Model $10;
	length Parameter $30;
	set PEforQ1CC PEforQ2CC PEforQ3CC PEforQ4CC PEforAllCC indsname=M;
	*THisISM=M;
	where Parameter="Intercept" or Parameter="lnEnplanement"
	or Parameter="lnIncome" or Parameter="lnPopulation";

	if	M="WORK.PEFORQ1CC" then Model="Model1";
		else if M="WORK.PEFORQ2CC" then Model="Model2";
		else if M="WORK.PEFORQ3CC" then Model="Model3";
		else if M="WORK.PEFORQ4CC" then Model="Model4";
		else if M="WORK.PEFORALLCC" then Model="Model5";

	if Probt le 0.01 then Star="***";
		else if Probt le 0.05 then Star="**";
		else if Probt le 0.1 then Star="*";

	EditedResults=cats(Put(Estimate,comma16.5),star);
	output;

	EditedResults=cats("(",put(StdErr,comma16.5),")");
	output;

run;



proc sort data=Table_LongCC out=Table_Long_SortedCC;
	by Model Parameter;
run;


/* Step 2: Create separate results columns for each model */
data Model1Results(rename=(EditedResults=Model1))
	Model2Results(rename=(EditedResults=Model2))
	Model3Results(rename=(EditedResults=Model3))
	Model4Results(rename=(EditedResults=Model4))
	Model5Results(rename=(EditedResults=Model5));
	set Table_Long_SortedCC;

	if Model="Model1" then output Model1Results;
		else if Model="Model2" then output Model2Results;
		else if Model="Model3" then output Model3Results;
```

```
		else if Model="Model4" then output Model4Results;
		else if Model="Model5" then output Model5Results;
	drop Model;
	keep Parameter EditedResults;
run;

data Table_WideCC;
	merge Model1Results Model2Results Model3Results Model4Results Model5Results;
	by Parameter;

	if Parameter="lnEnplanement" then Order=1;
		else if Parameter="lnPopulation" then Order=2;
		else if Parameter="lnIncome" then Order=3;
		else if substr(Parameter,1,9)="Intercept" then Order=4;

	if mod(_n_,2)=1 then Regressors=Parameter;

run;


/* Order the variables in the results table */
proc sort data=Table_WideCC out=Table_Wide_SortedCC(drop=Order Parameter);
	by Order;
run;


/* Step 4: Create the rows for other statistics */
/* County and Year Fixed Effects */
Data ControlCC;
	Regressors="County";
	Model1="Yes";
	Model2="Yes";
	Model3="Yes";
	Model4="Yes";
	Model5="Yes";

	output;
	Regressors="Year";
	output;

run;


/* The row for the number of observations */
Data NumofObsCC;
	merge ObsQ1CC(rename=(Nvalue1=NVModel1) drop=CValue1)
```

```
                    ObsQ2CC(rename=(Nvalue1=NVModel2) drop=CValue1)
                    ObsQ3CC(rename=(Nvalue1=NVModel3) drop=CValue1)
                    ObsQ4CC(rename=(Nvalue1=NVModel4) drop=CValue1)
                    ObsAllCC(rename=(Nvalue1=NVModel5) drop=CValue1);
        where Label1="Number of Observations";
        Model1=put(NVModel1,comma16.);
        Model2=put(NVModel2,comma16.);
        Model3=put(NVModel3,comma16.);
        Model4=Put(NVModel4,comma16.);
        Model5=Put(NVModel5,comma16.);

        keep Label1 Model1 Model2 Model3 Model4 Model5;
run;

/* The row for the adjusted R-Squared */
Data AdjRsqCC;
        merge AdjRsqQ1CC(rename=(cvalue1=Model1) drop=nvalue1)
                AdjRsqQ2CC(rename=(cvalue1=Model2) drop=nvalue1)
                AdjRsqQ3CC(rename=(cvalue1=Model3) drop=nvalue1)
                AdjRsqQ4CC(rename=(cvalue1=Model4) drop=nvalue1)
                AdjRsqAllCC(rename=(cvalue1=Model5) drop=nvalue1);
        Where Label1="Adjusted R-Square";
run;

/* The row for the F-test related to the Overall Significance of the model */
Data OSM1(rename=(EditedValue=Model1)) OSM2(rename=(EditedValue=Model2))
OSM3(rename=(EditedValue=Model3))
                OSM4(rename=(EditedValue=Model4)) OSM5(rename=(EditedValue=Model5));
        set OverallSigQ1CC OverallSigQ2CC OverallSigQ3CC OverallSigQ4CC
OverallSigAllCC indsname=M;
        where Effect="Model";
        if ProbF le 0.01 then Star="***";
                else if ProbF le 0.05 then Star="**";
                else if ProbF le 0.1 then Star="*";
        ThisIsM=M;

        Label1="Overall Significance";
        *EditedValue=cats(put(FValue,BESTw.2),Star);
        EditedValue=cats(put(FValue,e9.),Star);

                if M="WORK.OVERALLSIGQ1CC" then output OSM1;
                else if M="WORK.OVERALLSIGQ2CC" then output OSM2;
                else if M="WORK.OVERALLSIGQ3CC" then output OSM3;
                else if M="WORK.OVERALLSIGQ4CC" then output OSM4;
                else if M="WORK.OVERALLSIGALLCC" then output OSM5;
```

```
        keep Label1 EditedValue;
run;


Data OverallSigCC;
        merge OSM1 OSM2 OSM3 OSM4 OSM5;
        by Label1;
run;



/* Combine all rows for other statistics */
data OtherStatCC;
        set NumofObsCC AdjRsqCC OverallSigCC;
        rename Label1=Regressors;
run;

/* Add rows for other statistics to the table */
Data Table_Wide_Sorted_withStatCC;
        set Table_Wide_SortedCC ControlCC OtherStatCC;
run;



/* Print the clean results table */
/* New Code: The name of the excel file, the title of the results table, and its footnote are
modified */
ods excel file="/home/u63024456/MySAS/AQI Data
Analysis/FinalTableGDEnplanementQUARTILEScontrol.xlsx" options(Embedded_Titles="ON"
Embedded_Footnotes="ON"); /*Use the path to your MySAS folder */
Title "Table 4: Two-Way-Fixed-Effects Approach for Quartiles 1 through 4 with Control";
footnote justify=left "Source: EPA (2024), FAA (2024), with own calculations.";
footnote2 justify=left "Notes: Robust standard Errors are in Parentheses. *, **, *** indicate 10%,
5%,
        and 1% significance levels, respectively and are clustered at county levels. The data is
        unbalanced as the number of counties in each state in a given year are not constant over
time,
        as counties are added to the data or begin measuring AQI. Quartile 1=Model 1, etc.";
proc print data=Table_Wide_Sorted_withStatCC noobs;
        var Regressors;

        var Model1 Model2 Model3 Model4 Model5/ style(header)={Just=Center}
style(data)={Just=Center tagattr="type:string"};

        format Regressors $VariableName.;
run;
```

```
ods excel close;






/*****************************************************************
******************************************************************
                    MODELS 1-10 FOR PERCENTILES
******************************************************************
*****************************************************************/

proc means data=Combined2 P10 P20 P30 P40 P50 P60 P70 P80 P90;
	var id Agg_CY_Enplanement;
run;


/* (A 90th percentile) enplanements on GD */
ods output ParameterEstimates=PEforGDACCC DataSummary=ObsGDACCC
			FitStatistics=AdjRsqGDACCC Effects=OverallSigGDACCC;
Proc SurveyReg Data=CombinedControlFinal;
	*where (Agg_CY_Enplanement>17274506);
	where (id>7775);
	Class County_State Year / ref=First;
	ModelA: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
	*cluster County_State;
run;

/* (B 80th percentile) enplanement on GD */
ods output ParameterEstimates=PEforGDBCCC DataSummary=ObsGDBCCC
			FitStatistics=AdjRsqGDBCCC Effects=OverallSigGDBCCC;
Proc SurveyReg Data=CombinedControlFinal;
	*where (17274506>=Agg_CY_Enplanement>5054191);
	where (7775>=id>6911);
	Class County_State Year / ref=First;
	ModelB: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
	*cluster County_State;
run;


/* (C 70th percentile) Enplanement on GD */
```

```
ods output ParameterEstimates=PEforGDCCCC DataSummary=ObsGDCCCC
                        FitStatistics=AdjRsqGDCCCC Effects=OverallSigGDCCCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (5054191>=Agg_CY_Enplanement>2128405);
        where (6911>=id>6047);
        Class County_State Year / ref=First;
        ModelC: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
run;

/* (D 60th percentile) Enplanement on GD*/
ods output ParameterEstimates=PEforGDDCCC DataSummary=ObsGDDCCC
                        FitStatistics=AdjRsqGDDCCC Effects=OverallSigGDDCCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (2128405>=Agg_CY_Enplanement>846369);
        where (6047>=id>5183);
        Class County_State Year / ref=First;
        ModelD: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
run;

/* (E 50th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDECCC DataSummary=ObsGDECCC
                        FitStatistics=AdjRsqGDECCC Effects=OverallSigGDECCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (846369>=Agg_CY_Enplanement>429523);
        where (5183>=id>4319.5);
        Class County_State Year / ref=First;
        ModelE: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
run;


/* (F 40th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDFCCC DataSummary=ObsGDFCCC
                        FitStatistics=AdjRsqGDFCCC Effects=OverallSigGDFCCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (429523>=Agg_CY_Enplanement>263217);
        where (4319.5>=id>3456);
        Class County_State Year / ref=First;
        ModelF: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
```

```
run;

/* (G 30th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDGCCC DataSummary=ObsGDGCCC
                        FitStatistics=AdjRsqGDGCCC Effects=OverallSigGDGCCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (263217>=Agg_CY_Enplanement>138423);
        where (3456>=id>2592);
        Class County_State Year / ref=First;
        ModelG: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
run;

/* (H 20th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDHCCC DataSummary=ObsGDHCCC
                        FitStatistics=AdjRsqGDHCCC Effects=OverallSigGDHCCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (138423>=Agg_CY_Enplanement>54728);
        where (2592>=id>1728);
        Class County_State Year / ref=First;
        ModelH: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
run;

/* (I 20th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDICCC DataSummary=ObsGDICCC
                        FitStatistics=AdjRsqGDICCC Effects=OverallSigGDICCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (54728>=Agg_CY_Enplanement>23240);
        where (1728>=id>864);
        Class County_State Year / ref=First;
        ModelI: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
        *cluster County_State;
run;

/* (J 10th percentile) Enplanement on GD */
ods output ParameterEstimates=PEforGDJCCC DataSummary=ObsGDJCCC
                        FitStatistics=AdjRsqGDJCCC Effects=OverallSigGDJCCC;
Proc SurveyReg Data=CombinedControlFinal;
        *where (23240>=Agg_CY_Enplanement);
        where (864>=id);
        Class County_State Year / ref=First;
```

```
	ModelJ: Model lnGD = lnEnplanement lnIncome lnPopulation County_State Year /
Solution AdjRsq;
	*cluster County_State;
run;



/*****************************************************************
*****************************************************************
                              BUILD RESULTS FOR PERCENTILES
*****************************************************************
*****************************************************************/


/* BUILD RESULTS */

Data Table_LongCCC;
	length Model $10;
	length Parameter $30;
	set PEforGDACCC PEforGDBCCC PEforGDCCCC PEforGDDCCC PEforGDECCC
PEforGDFCCC PEforGDGCCC PEforGDHCCC PEforGDICCC PEforGDJCCC indsname=M;
	*THisISM=M;
	where Parameter="Intercept" or Parameter="lnEnplanement"
	or Parameter="lnIncome" or Parameter="lnPopulation";


	if	M="WORK.PEFORGDACCC" then Model="Model10";
		else if M="WORK.PEFORGDBCCC" then Model="Model9";
		else if M="WORK.PEFORGDCCCC" then Model="Model8";
		else if M="WORK.PEFORGDDCCC" then Model="Model7";
		else if M="WORK.PEFORGDECCC" then Model="Model6";
		else if M="WORK.PEFORGDFCCC" then Model="Model5";
		else if M="WORK.PEFORGDGCCC" then Model="Model4";
		else if M="WORK.PEFORGDHCCC" then Model="Model3";
		else if M="WORK.PEFORGDICCC" then Model="Model2";
		else if M="WORK.PEFORGDJCCC" then Model="Model1";

	if Probt le 0.01 then Star="***";
		else if Probt le 0.05 then Star="**";
		else if Probt le 0.1 then Star="*";

	EditedResults=cats(Put(Estimate,comma16.5),star);
	output;

	EditedResults=cats("(",put(StdErr,comma16.5),")");
	output;
```

```
run;


proc sort data=Table_LongCCC out=Table_Long_SortedCCC;
	by Model Parameter;
run;


/* Step 2: Create separate results columns for each model */
data Model1Results(rename=(EditedResults=Model1))
	Model2Results(rename=(EditedResults=Model2))
	Model3Results(rename=(EditedResults=Model3))
	Model4Results(rename=(EditedResults=Model4))
	Model5Results(rename=(EditedResults=Model5))
	Model6Results(rename=(EditedResults=Model6))
	Model7Results(rename=(EditedResults=Model7))
	Model8Results(rename=(EditedResults=Model8))
	Model9Results(rename=(EditedResults=Model9))
	Model10Results(rename=(EditedResults=Model10));
	set Table_Long_SortedCCC;

	if Model="Model1" then output Model1Results;
		else if Model="Model2" then output Model2Results;
		else if Model="Model3" then output Model3Results;
		else if Model="Model4" then output Model4Results;
		else if Model="Model5" then output Model5Results;
		else if Model="Model6" then output Model6Results;
		else if Model="Model7" then output Model7Results;
		else if Model="Model8" then output Model8Results;
		else if Model="Model9" then output Model9Results;
		else if Model="Model10" then output Model10Results;
	drop Model;
	keep Parameter EditedResults;
run;

data Table_WideCCC;
	merge Model1Results Model2Results Model3Results Model4Results Model5Results
Model6Results
	Model7Results Model8Results Model9Results Model10Results;
	by Parameter;

	if Parameter="lnEnplanement" then Order=1;
		else if Parameter="lnPopulation" then Order=2;
		else if Parameter="lnIncome" then Order=3;
```

```
		else if substr(Parameter,1,9)="Intercept" then Order=4;

	*if Parameter="DID" then Order=1;
		*else if substr(Parameter,1,9)="Intercept" then Order=2;

	if mod(_n_,2)=1 then Regressors=Parameter;

run;


/* Order the variables in the results table */
proc sort data=Table_WideCCC out=Table_Wide_SortedCCC(drop=Order Parameter);
	by Order;
run;


/* Step 4: Create the rows for other statistics */
/* County and Year Fixed Effects */
Data ControlCCC;
	Regressors="County";
	Model1="Yes";
	Model2="Yes";
	Model3="Yes";
	Model4="Yes";
	Model5="Yes";
	Model6="Yes";
	Model7="Yes";
	Model8="Yes";
	Model9="Yes";
	Model10="Yes";
	output;
	Regressors="Year";
	output;

run;


/* The row for the number of observations */
Data NumofObsCCC;
	merge ObsGDACCC(rename=(Nvalue1=NVModel10) drop=CValue1)
			ObsGDBCCC(rename=(Nvalue1=NVModel9) drop=CValue1)
			ObsGDCCCC(rename=(Nvalue1=NVModel8) drop=CValue1)
			ObsGDDCCC(rename=(Nvalue1=NVModel7) drop=CValue1)
			ObsGDECCC(rename=(Nvalue1=NVModel6) drop=CValue1)
			ObsGDFCCC(rename=(Nvalue1=NVModel5) drop=CValue1)
			ObsGDGCCC(rename=(Nvalue1=NVModel4) drop=CValue1)
```

```
                    ObsGDHCCC(rename=(Nvalue1=NVModel3) drop=CValue1)
                    ObsGDICCC(rename=(Nvalue1=NVModel2) drop=CValue1)
                    ObsGDJCCC(rename=(Nvalue1=NVModel1) drop=CValue1);
        where Label1="Number of Observations";
        Model1=put(NVModel1,comma16.);
        Model2=put(NVModel2,comma16.);
        Model3=put(NVModel3,comma16.);
        Model4=Put(NVModel4,comma16.);
        Model5=put(NVModel5,comma16.);
        Model6=put(NVModel6,comma16.);
        Model7=put(NVModel7,comma16.);
        Model8=put(NVModel8,comma16.);
        Model9=put(NVModel9,comma16.);
        Model10=put(NVModel10,comma16.);
        keep Label1 Model1 Model2 Model3 Model4 Model5 Model6 Model7 Model8 Model9
Model10;
run;

/* The row for the adjusted R-Squared */
Data AdjRsqCCC;
        merge AdjRsqGDACCC(rename=(cvalue1=Model10) drop=nvalue1)
                AdjRsqGDBCCC(rename=(cvalue1=Model9) drop=nvalue1)
                AdjRsqGDCCCC(rename=(cvalue1=Model8) drop=nvalue1)
                AdjRsqGDDCCC(rename=(cvalue1=Model7) drop=nvalue1)
                AdjRsqGDECCC(rename=(cvalue1=Model6) drop=nvalue1)
                AdjRsqGDFCCC(rename=(cvalue1=Model5) drop=nvalue1)
                AdjRsqGDGCCC(rename=(cvalue1=Model4) drop=nvalue1)
                AdjRsqGDHCCC(rename=(cvalue1=Model3) drop=nvalue1)
                AdjRsqGDICCC(rename=(cvalue1=Model2) drop=nvalue1)
                AdjRsqGDJCCC(rename=(cvalue1=Model1) drop=nvalue1);
        Where Label1="Adjusted R-Square";
run;

/* The row for the F-test related to the Overall Significance of the model */
Data OSM1(rename=(EditedValue=Model1)) OSM2(rename=(EditedValue=Model2))
OSM3(rename=(EditedValue=Model3))
                OSM4(rename=(EditedValue=Model4)) OSM5(rename=(EditedValue=Model5))
OSM6(rename=(EditedValue=Model6))
                OSM7(rename=(EditedValue=Model7)) OSM8(rename=(EditedValue=Model8))
OSM9(rename=(EditedValue=Model9))
                OSM10(rename=(EditedValue=Model10));
        set OverallSigGDACCC OverallSigGDBCCC OverallSigGDCCCC
OverallSigGDDCCC OverallSigGDECCC OverallSigGDFCCC
        OverallSigGDGCCC OverallSigGDHCCC OverallSigGDICCC OverallSigGDJCCC
indsname=M;
        where Effect="Model";
```

```
        if ProbF le 0.01 then Star="***";
            else if ProbF le 0.05 then Star="**";
            else if ProbF le 0.1 then Star="*";
        ThisIsM=M;

        Label1="Overall Significance";
        *EditedValue=cats(put(FValue,BESTw.2),Star);
        EditedValue=cats(put(FValue,e9.),Star);

            if M="WORK.OVERALLSIGGDACCC" then output OSM10;
            else if M="WORK.OVERALLSIGGDBCCC" then output OSM9;
            else if M="WORK.OVERALLSIGGDCCCC" then output OSM8;
            else if M="WORK.OVERALLSIGGDDCCC" then output OSM7;
            else if M="WORK.OVERALLSIGGDECCC" then output OSM6;
            else if M="WORK.OVERALLSIGGDFCCC" then output OSM5;
            else if M="WORK.OVERALLSIGGDGCCC" then output OSM4;
            else if M="WORK.OVERALLSIGGDHCCC" then output OSM3;
            else if M="WORK.OVERALLSIGGDICCC" then output OSM2;
            else if M="WORK.OVERALLSIGGDJCCC" then output OSM1;
        keep Label1 EditedValue;
run;


Data OverallSigCCC;
        merge OSM1 OSM2 OSM3 OSM4 OSM5 OSM6 OSM7 OSM8 OSM9 OSM10;
        by Label1;
run;



data OtherStatCCC;
        set NumofObsCCC AdjRsqCCC OverallSigCCC;
        rename Label1=Regressors;
run;



/* Add rows for other statistics to the table */
Data Table_Wide_Sorted_withStatCCC;
        set Table_Wide_SortedCCC ControlCCC OtherStatCCC;
run;
```

```
/* Print the clean results table */
/* New Code: The name of the excel file, the title of the results table, and its footnote are
modified */
ods excel file="/home/u63024456/MySAS/AQI Data
Analysis/FinalTableGDEnplanementPERCENTILEScontrol.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON"); /*Use the path to your MySAS
folder */
Title "Table 5: Two-Way-Fixed-Effects Approach for 10 Percentiles with Control";
footnote justify=left "Source: EPA (2024), FAA (2024), with own calculations.";
footnote2 justify=left "Notes: Robust standard Errors are in Parentheses.
                              *, **, *** indicate 10%, 5%, and 1% significance levels, respectively and
are clustered at county levels.
                              The data is unbalanced as the number of counties in each state in a given
year are not constant over time,
                              as counties are added to the data or begin measuring AQI.";
proc print data=Table_Wide_Sorted_withStatCCC noobs;
          var Regressors;

          var Model1 Model2 Model3 Model4 Model5 Model6 Model7 Model8 Model9 Model10 /
style(header)={Just=Center} style(data)={Just=Center tagattr="type:string"};

          format Regressors $VariableName.;
run;
ods excel close;
```